



Universidade Nova de Lisboa
Instituto de Higiene e Medicina Tropical

Uma abordagem metabolómica no diagnóstico da tuberculose

Ricardo Diogo Alves Conde

DISSERTAÇÃO PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS BIOMÉDICAS
ESPECIALIDADE EM BIOLOGIA MOLECULAR EM MEDICINA TROPICAL E
INTERNACIONAL

OUTUBRO DE 2016



Universidade Nova de Lisboa
Instituto de Higiene e Medicina Tropical

Uma abordagem metabolómica no diagnóstico da tuberculose

Autor: Ricardo Diogo Alves Conde

Orientador: Doutora Ana Varela Coelho

Coorientador: Doutor Manolis Matzapetakis

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de
Mestre em Ciências Biomédica

A presente dissertação foi redigida de acordo com o novo acordo ortográfico

Dedicatória

Para quem acredita

Agradecimentos

Gostaria de começar a agradecer à minha orientadora, Ana Coelho, e ao meu coorientador, Manolis Matzapetakis, por esta oportunidade única de trabalhar numa área até então desconhecida por mim. A eles quero também agradecer por me transmitirem os seus conhecimentos nas suas áreas respetivas e por me ajudarem durante este ano que passou. Quero também agradecer o facto de me terem proporcionado a participação em cursos de NMR e de proteómica.

Não poderia de deixar de agradecer aos meus colegas de laboratório, principalmente à Mariana Palma, por me ter esclarecido qualquer duvida que tive. Quero agradecer também à Mariana pela companhia no almoço e pelos conselhos que me deu ao longo deste ano.

Quero agradecer aos CDPs de Vendas-Novas e de Almada-Seixal, por realizarem as recolhas de sangue dos indivíduos que participaram neste estudo. Quero também agradecer ao INSA, pela realização dos ensaios IGRA e pela descontaminação das amostras.

Aos meus amigos, agradeço por me terem “aturado” durante esta minha jornada. Quero agradecer especialmente ao Fred, ao Francisco, à Lúcia, à Cátia, à Inês, ao Ricardo, à Liliana, ao Samuel, ao Miguel. Obrigado por serem os amigos que são e por acreditarem em mim.

Como não poderia deixar passar, quero agradecer o apoio da minha família, principalmente dos meus pais, que muito se esforçaram para eu realizar esta jornada.

Resumo

Este trabalho teve como foco principal a identificação de potenciais biomarcadores para o diagnóstico da tuberculose, utilizando uma abordagem metabolômica. Estudos de metabolômica utilizando amostras de biofluidos são hoje em dia muito comuns na descoberta de biomarcadores para diagnóstico de doenças, além de auxiliarem na compreensão dos seus mecanismos patológicos. A tuberculose tornou-se uma doença emergente a nível mundial, sendo uma das principais causas de morte. *Mycobacterium tuberculosis* é o agente patogénico responsável pela tuberculose.

Neste sentido, foram adquiridos espectros de ^1H NMR de amostras de soro sanguíneo de indivíduos saudáveis e de pacientes com tuberculose, utilizando um espectrómetro de NMR de 800 MHz. A aquisição de espectros foi realizada após a otimização do protocolo de preparação de amostras de soro. Para isso, verificou-se qual a melhor sequência de pulsos a ser utilizada (CPMG ou NOESY), qual o efeito da ultrafiltração no perfil dos espectros e determinou-se a estabilidade da amostra. A abordagem metabolômica utilizada neste estudo envolveu uma análise do perfil metabólico e uma análise *fingerprinting* das amostras de soro. A identificação e quantificação dos metabolitos foi efetuada com o software Chenomx NMR Suite 8.1. A identificação foi auxiliada por bases de dados e espectros de NMR bidimensionais. Após a identificação/quantificação dos metabolitos procedeu-se à análise estatística. Para os dados obtidos a partir da análise do perfil metabólico foi realizada uma análise uni e multivariada dos dados. No caso da análise *fingerprinting* apenas se procedeu à realização da análise multivariada.

Os resultados das duas análises foram similares, identificando os mesmos metabolitos responsáveis pela discriminação dos grupos experimentais (controles e doentes). Foram encontrados em maiores concentrações no soro de indivíduos com tuberculose intermediários de nucleosídeos (hipoxantina e inosina), glícidos (manose e glucose), aminoácidos (ornitina, glutamato, serina e aspartato) e ácidos orgânicos (lactato). Já aminoácidos como a asparagina, a metionina, a cisteína, a lisina e a π -metilhistidina, e ácidos orgânicos, como o 3-hidroxisobutirato e o citrato, foram encontrados em menores concentrações no soro dos indivíduos com tuberculose.

A partir destes resultados verificou-se que os principais processos alterados com reflexo nos níveis metabólicos do soro de doentes, estavam relacionados com a biossíntese de proteínas, com a reciclagem de amónia, com o metabolismo do aspartato e com o metabolismo da metionina. Os resultados obtidos neste estudo corroboram outros estudos já publicados

Palavras-chave: tuberculose; *Mycobacterium tuberculosis*; metabolômica; NMR; análise multivariada

Abstract

This work had as major focus the identification of potential biomarkers for tuberculosis diagnosis, using a metabolomic approach. Metabolomics studies using biofluids samples are nowadays very commons in the discovery of biomarkers for diseases diagnosis, in addition to assist in the compression of their mechanisms of pathogenesis. Tuberculosis has become an emerging disease worldwide and is a major cause of death. *Mycobacterium tuberculosis* is the pathogen responsible for tuberculosis.

In this study were acquired ^1H NMR spectra of blood serum samples of healthy subjects and of patients with tuberculosis, using a 800 MHz NMR spectrometer. The spectra acquisition was performed after optimization of serum samples preparation protocol. For this, it was found that the better spectra type to be used is NOESY and were evaluated the effect of ultrafiltration on the spectra profile and determined the sample stability.

The metabolomics approach used in this study involves a metabolic profiling and a fingerprinting analysis of the obtained NMR data. The identification and quantification of metabolites was performed with Chenomx NMR Suite 8.1 software. The identification was assist using databases and two-dimensional NMR spectra. After metabolites identification and quantification, we proceeded to the statistical analysis. For the results obtained with the metabolic profiling method both uni- and multivariate analysis were performed. On behalf of the fingerprinting method results were only evaluated by multivariate analysis

The results of both analysis were similar, identifying the same metabolites responsible for discrimination between experimental groups (controls and patients). In serum of TB-infected individuals were found in higher concentrations nucleosides intermediaries (hypoxanthine and inosine), carbohydrates (mannose and glucose), amino acids (ornithine, glutamate, serine and aspartate) and organic acids (lactate). Amino acids as asparagine, methionine, cysteine, lysine and π -methylhistidine and organic acids such as 3-hydroxyisobutyrate and citrate were found at lower concentrations in the serum of patients.

From these results, it was found that the major metabolic changes reflected in the serum of patients were related with protein biosynthesis, ammonia recycling aspartate and methionine metabolism. These results support other studies already published.

Keywords: tuberculosis; *Mycobacterium tuberculosis*; metabolomics; NMR; multivariate analysis

Índice Geral

Dedicatória.....	v
Agradecimentos	vii
Resumo	ix
Abstract.....	xi
Índice Geral	xiii
Índice de Figuras.....	xvii
Índice de Tabelas	xiii
Lista de abreviaturas	xiii
Capítulo I: Introdução.....	1
1.1 Tuberculose, uma doença emergente	3
1.2 O género Mycobacterium	4
1.2.1 Mycobacterium tuberculosis	5
1.3 Patogénese e Imunologia.....	6
1.4 Diagnóstico.....	8
1.5 Tratamento	11
1.6 O sangue, um biofluido de excelência	12
1.6.1 Constituição sanguínea	12
1.6.2 Plasma versus soro	12
1.6.3 Metaboloma do soro.....	13
1.7 Metabolómica	14
1.8 Ressonância magnética nuclear (NMR)	16
1.8.1 NMR na metabolómica	19
1.8.2 Preparação de amostras para NMR	20

1.9	Análise estatística.....	22
1.9.1	Análise univariada	22
1.9.2	Análise multivariada.....	22
1.10	. Biomarcadores no diagnóstico da tuberculose.....	26
Capítulo II: Objetivos		29
Capítulo III: Materiais e Métodos.....		33
3.1	Metodologia.....	35
3.2	Caracterização do perfil dos voluntários	36
3.3	Recolha das amostras de soro sanguíneo.....	37
3.4	Preparação das amostras para NMR	38
3.5	Aquisição de espectros de NMR	39
3.6	Processamento dos espectros de NMR	40
3.7	Identificação e quantificação de metabolitos.....	41
3.8	Binning	41
3.9	Processamento dos dados para análise estatística	42
3.10	Análise Univariada	43
3.11	Análise Multivariada	43
3.12	Validação dos modelos de análise multivariada.....	44
3.13	Análise das vias metabólicas.....	44
Capítulo IV: Resultados e Discussão.....		45
4.1	Caracterização dos voluntários.....	47
4.2	Otimização do protocolo para aquisição de espectros de NMR	50
4.3	Desconvolução dos espectros de ¹ H NMR de amostras de soro de indivíduos saudáveis e doentes.....	52
4.4	Análise Univariada	59
4.5	Análise multivariada	61
4.5.1	Seleção da dimensão dos bins	61

4.5.2 Remoção de outliers	63
4.5.3 Análise de variáveis externas.....	65
4.5.4 Discriminação entre amostras de indivíduos saudáveis e de indivíduos com tuberculose	69
4.6 Análise das vias metabólicas.....	75
Capítulo V: Conclusões	81
Capítulo VI: Perspetivas futuras	85
Capítulo VII: Bibliografia.....	89
Anexo I	103
Anexo II	107
Anexo III.....	109
Anexo IV	110
Anexo V: Gráficos de bigodes (análise univariada)	112
Anexo VI	115
Anexo VII	116
Anexo VIII.....	118
Anexo IX	120

Índice de Figuras

Figura 1: Taxa de incidência estimada da tuberculose em 2014 (Lewandowski et al. 2015).....	3
Figura 2: Evolução da taxa de notificação e de incidência de tuberculose em Portugal, 2004-2014 (Diniz et al. 2015)	4
Figura 3: Representação da parede celular de Mtb. 1. Bicamada fosfolipídica com proteínas integradas; 2. Peptidoglicanos; 3. Arabinogalactano; 4. Ácidos micólicos; 5. Porinas externas; 6. Lipoarabinomanano (LAM); 7. Lípidos e glicolípidos; 8. Proteínas de superfície (Barroso et al. 2014).	5
Figura 4: Estrutura química dos antibacilares de primeira linha utilizados no tratamento da tuberculose. A: Isoniazida; B: Rifampicina; C: Pirazinamida; D: Etambutol.....	11
Figura 5: Gráfico representativo do número de artigos publicados com os respectivos termos ("genomics", "transcriptomics", "proteomics" e "metabolomics") desde 2000 até 2015. A pesquisa do número de publicações em cada área foi realizada através do site ScienceDirect (www.sciencedirect.com).....	14
Figura 6: Componentes de um espectrómetro de ressonância magnética nuclear	18
Figura 7: Gráfico representativo do número de Publicações por ano (2001-2015) na área da metabolômica usando a técnica de NMR. Web of Science Scopus foi usado para pesquisar por este tipo de publicações usando os termos "metabolomics" e "NMR" em "All fields".....	19
Figura 8: Gráfico representativo do número de publicações no campo da metabolômica por ano (linha vermelha) contra o número de publicações que incluem o PLS-DA como uma ferramenta na análise de dados de metabolômica (linha azul) (de Thomas Reuters' ISI Web of Science1 usando as palavras chaves "metabolomics", "PLS" e "PLS-DA" (adaptado de 82).....	24

Figura 9: Esquema da metodologia utilizada para este trabalho. 1: recolha de amostras de sangue; 2: separação do soro; ensaio IGRA; 3: remoção das micobactérias; 4: armazenamento a -80°C; 5: transporte das amostras para o ITQB; 6: armazenamento a -80°C; 7: ultrafiltração das amostras de soro; 8: adição de buffer e D2O; 9: aquisição de espectros; 10: Identificação e quantificação de metabolitos; 11: análise univariada; 12: Análise multivariada; 13: MSEA 35

Figura 10: Histograma representativo da distribuição dos indivíduos saudáveis, com tuberculose pulmonar e com tuberculose extrapulmonar por género. 47

Figura 11: Histograma representativo da distribuição dos indivíduos saudáveis, com tuberculose pulmonar e com tuberculose extrapulmonar pela idade. < 30 anos; 30 a 50 anos; > 50 anos..... 48

Figura 12: Histograma representativo da distribuição dos indivíduos saudáveis, com tuberculose pulmonar e com tuberculose extrapulmonar segundo o seu IMC. L: IMC baixo; N: IMC normal; H: IMC elevado..... 49

Figura 13: Histograma representativo da distribuição dos indivíduos saudáveis, com tuberculose pulmonar e com tuberculose extrapulmonar segundo os seus hábitos tabágicos. S: fumador; N: não fumador; EX: ex-fumador 49

Figura 14: Representação dos espectros NOESY e CPMG de uma amostra controlo. A: espectro 1D NOESY sem prévia ultrafiltração da amostra; B: Espectro CPMG sem prévia ultrafiltração da amostra; C: Espectro CPMG com prévia ultrafiltração da amostra; D: Espectro 1D NOESY com prévia ultrafiltração da amostra..... 51

Figura 15: Espectro 1D NOESY de uma amostra do grupo controlo, adquirido com prévia ultrafiltração da amostra. As regiões A a G apresentadas, estão com mais detalhe nas figuras A1-A7 do anexo I.DSS..... 54

Figura 16: Histograma ilustrativo da comparação das concentrações metabólicas quantificadas por NMR, em amostras de indivíduos saudáveis, e descritas na literatura, relativamente aos metabolitos 1 e 2. 56

Figura 17: Histograma ilustrativo da comparação das concentrações metabólicas quantificadas por NMR, em amostras de indivíduos saudáveis, e descritas na literatura, relativamente aos metabolitos 3 a 22	56
Figura 18: Histograma ilustrativo da comparação das concentrações metabólicas quantificadas por NMR, em amostras de indivíduos saudáveis, e descritas na literatura, relativamente aos metabolitos 23 a 34 e 36 a 53.....	57
Figura 19: Teste t-Sudent. As variáveis representadas acima da linha obtiveram um valor de p inferior a 0.05	60
Figura 20: Gráfico de scores do modelo de PCA utilizado para a exclusão de outliers, o caso da análise fingerprinting	64
Figura 21: Gráfico de scores do modelo de PCA utilizado para a exclusão de outliers, o caso da análise do perfil metabólico	65
Figura 22: Gráficos de scores do modelo de PCA referentes à análise fingerprinting (A) e à análise do perfil metabólico (B), usado para comparar amostras do grupo controlo e do grupo de doentes.A	Erro! Marcador não definido.
Figura 23: Gráfico de scores do modelo de PLS-DA para a análise fingerprinting.....	Erro! Marcador não definido.
Figura 24: Gráfico de scores do modelo de PLS-DA para a análise do perfil metabólico.....	72
Figura 25: Gráfico de loadings do modelo de PLS-DA para a análise do perfil metabólico As variáveis marcadas a vermelho e a azul encontram-se em menores e maiores concentrações, respetivamente, no soro de indivíduos com tuberculose.....	73
Figura 26: Gráfico de loadings do modelo de PLS-DA para a análise fingerprinping. Os pontos a azul representam a tendência do conjunto de amostras de cada grupo.....	74

Figura 27: Análise de enriquecimento do conjunto de metabolitos representativa dos processos metabólicos principais alterados no soro de indivíduos com tuberculose ativa.....	76
Figura 28: Metabolismo da metionina 1: metionina adenosiltransferase. A vermelho e a azul estão marcados os metabolitos com níveis elevados e reduzidos, respetivamente no soro de doentes.....	77
Figura A 1: Região espectral A, de uma amostra controlo, entre os 0.8 e os 1.8 ppm. 2: lactato; 4: alanina; 6: valina; 8: lisina; 12: leucina; 15: isoleucina; 21: 3-hidroxi-butirato; 26: 2-hidroxi-butirato; 35: fucose; 37: 2-aminobutirato; 46: 2-hidroxiisovalerato; 47: propilenoglicol; 59: etanol; 61: DSS. Ampliação 64x	103
Figura A 2: Região B entre os 1.8 e os 2.8 ppm de um espectro de uma amostra controlo. 3: glutamina; 6: valina; 7: prolina; 8: lisina; 10: arginina; 13: citrato; 20: ornitina; 21: 3-hidroxi-butirato; 24: acetato; 27: glutamato; 33: piruvato; 34: metionina; 38: acetoacetato; 40: aspartato; 50: 2-oxoisocaproato; 57: acetona. Ampliação 64x	103
Figura A 3: Região C entre os 2.8 e os 3.5 ppm de um espectro de uma amostra controlo. 1: glucose; 7: prolina; 8: lisina; 14: creatinina; 17: histidina; 18: cistina; 20: ornitina; 23: asparagina; 29: betaína; 30: carnitina; 36: cisteína; 44: colina; 56: 1,5-anidro sorbitol; 58: cafeína; 61: DSS. Ampliação 16x	104
Figura A 4: Região D entre os 3.5 e os 4 ppm de um espectro de uma amostra controlo. 1: glucose; 4: alanina; 5: glicina; 6: valina; 9: treonina; 11: serina; 15: isoleucina; 19: tirosina; 25: manose; 29: betaína; 31: creatina; 32: mio-inositol; 56: 1,5-anidro sorbitol; 58: cafeína; 59: etanol; 62: glicerol Ampliação 8x	104
Figura A 5: Região E entre os 4 e os 4.7 ppm de um espectro de uma amostra controlo. 1: glucose; 2: lactato; 7: prolina; 9: treonina; 14: creatinina; 21: 3-hidroxi-butirato; 28: ascorbato; 35: fucose; 45: trans-4-hidroxi-L-prolina; 49: uridina; 54: trigonelina. Ampliação 32x	105
Figura A 6: Região F entre os 4,7 e os 6,5 ppm de um espectro de uma amostra controlo. 1: glucose; 25: manose; 60: ureia Ampliação 32x	105

Figura A 7: Região G entre os 6,5 e os 9 ppm de um espectro de uma amostra controle. 17: histidina; 19: tirosina; 22: fenilalanina; 42: formato; 43: hipoxantina; 48: π -metilhistidina; 49: uridina; 51: triptofano; 52: τ -metilhistidina; 53: inosina. Ampliação 32x	106
Figura B 1: região do espectro TOCSY utilizada para a identificação do 3-hidroxisobutirato.....	107
Figura B 2: região do espectro TOCSY utilizada para a identificação do 2-aminobutirato	107
Figura B 3: região do espectro TOCSY utilizada para a identificação da frutose	108
Figura B 4: região do espectro TOCSY utilizada para a identificação do propileno glicol.....	108
Figura C 1: região do espectro J-resolved de 0.74 a 1.14 ppm.....	109
Figura C 2: região do espectro J-resolved de 4.14 a 4.54 ppm.....	109
Figura D: Gráfico de scores do modelo de PLS-DA e respetivo teste de permutação utilizado para a escolha do tamanho ideal dos bins. Análises realizadas com bins de 0.01 ppm (A), 0.02 ppm (B), 0.03 ppm (C) e 0.04 ppm (D).....	115
Figura E 1 1: Gráficos de scores dos modelos de PLS-DA das amostras controle, utilizados para analisar as variáveis externas através da análise fingerprinting. A: género; B: idade; C: IMC; D: Hábitos tabágicos	116
Figura E 1 2: Testes de permutação dos modelos de PLS-DA das amostras controle, utilizados para analisar as variáveis externas através da análise fingerprinting. A: género; B: idade; C: IMC; D: Hábitos tabágicos. Os pontos verdes representam os valores de R^2 e os pontos azuis representam os valores de Q^2B	116
Figura E 1 3: Gráficos de scores dos modelos de PLS-DA das amostras controle, utilizados para analisar as variáveis externas através da análise do perfil metabólico. A: género; B: idade; C: IMC; D: Hábitos tabágicos.....	117

Figura E 1 4: Testes de permutação dos modelos de PLS-DA das amostras controlo, utilizados para analisar as variáveis externas através da análise do perfil metabólico. A: género; B: idade; C: IMC; D: Hábitos tabágicos. Os pontos verdes representam os valores de R^2 e os pontos azuis representam os valores de Q^2	117
Figura F 1: Gráficos de scores dos modelos de PLS-DA das amostras de doentes, utilizados para analisar as variáveis externas através da análise fingerprinting. A: género; B: idade; C: IMC; D: Hábitos tabágicos	117
Figura F 2: Testes de permutação dos modelos de PLS-DA das amostras de doentes, utilizados para analisar as variáveis externas através da análise fingerprinting. A: género; B: idade; C: IMC; D: Hábitos tabágicos. Os pontos verdes representam os valores de R^2 e os pontos azuis representam os valores de Q^2	118
Figura F 3: Gráficos de scores dos modelos de PLS-DA das amostras de doentes, utilizados para analisar as variáveis externas através da análise do perfil metabólico. A: género; B: idade; C: IMC; D: Hábitos tabágicos.....	119
Figura F 4: Testes de permutação dos modelos de PLS-DA das amostras de doentes, utilizados para analisar as variáveis externas através da análise fingerprinting. A: género; B: idade; C: IMC; D: Hábitos tabágicos. Os pontos verdes representam os valores de R^2 e os pontos azuis representam os valores de Q^2	119
Figura G 1: Testes de permutação dos modelos de PLS-DA das amostras de controlos e de doentes, utilizados para analisar as variáveis externas através da análise fingerprinting. Os pontos verdes representam os valores de R^2 e os pontos azuis representam os valores de Q^2	120
Figura G 2: Testes de permutação dos modelos de PLS-DA das amostras dos controlos e de doentes, utilizados para analisar as variáveis externas através da análise do perfil metabólico. Os pontos verdes representam os valores de R^2 e os pontos azuis representam os valores de Q^2 ..	120
Figura G 3: Representação dos VIPs para o modelo de PLS-DA utilizando os dados da análise fingerprinting	121

Índice de Tabelas

Tabela 1: Diferentes métodos de diagnósticos aplicados. PPV: valores preditivos positivos, NPV: valores preditivos negativos, MODS: suscetibilidade a drogas observada por microscopia, E-MTD: potenciador direto de Mycobacterium tuberculosis (du Preez & Loots 2014).....	9
Tabela 2: Número de casos com resistência aos antibacilares de primeira linha em Portugal em 2014 (Diniz et al. 2015).	11
Tabela 3: Parâmetros de aquisição de espectros 1D NOESY	39
Tabela 4: Parâmetros de aquisição de espectros bidimensionais	40
Tabela 5: Média e desvios padrão das concentrações metabólicas encontradas nas amostras do grupo controlo e nas amostras do grupo de doentes. Valores normais descritos na literatura para a concentração metabólica dos metabolitos identificados. NQ: não quantificável; NE: não se apresenta descrito na literatura. A numeração dos metabolitos, presentes na tabela, corresponde à numeração que vai ser descrita ao longo deste capítulo e à numeração presente nas figura A 1 – A7 do anexo I.	59
Tabela 6: metabolitos e respetivos valores de p do teste t-Sudent	60
Tabela 7: Parâmetros relativos aos modelos de PCA usado para avaliar o tamanho dos bins...	62
Tabela 8: Parâmetros relativos aos modelos de PLS-DA usado para avaliar o tamanho dos bins	63
Tabela 9: Parâmetros relativos aos modelos de PLS-DA usados para avaliar as variáveis externas nas amostras controlo para a análise do perfil metabólico	66
Tabela 10: Parâmetros relativos aos modelos de PLS-DA usados para avaliar as variáveis externas nas amostras controlo para a análise fingerprinting.....	67

Tabela 11: Parâmetros relativos aos modelos de PLS-DA usados para avaliar as variáveis externas nas amostras de doentes para a análise do perfil metabólico	68
Tabela 12: Parâmetros relativos aos modelos de PLS-DA usados para avaliar as variáveis externas nas amostras de doentes para a análise fingerprinting	68
Tabela 13: Parâmetros de validação dos modelos da análise multivariada para a comparação de amostras do grupo de doentes pulmonares e de doentes extrapulmonares	70
Tabela 14: Parâmetros de validação dos modelos de PCA	71
Tabela 15: Parâmetros de validação dos modelos de PLS-DA	72
Tabela 16: Metabolitos e respetivos sinais responsáveis pela separação entre as amostras do grupo controlo e as amostras do grupo de doentes utilizando a análise do perfil metabólico e a análise fingerprinting	Erro! Marcador não definido.
Tabela 17:. Principais processos metabólicos alterados, como respetivo número de metabolitos, valores de p e taxa de descoberta de falsos positivos (FDR)	76
Tabela 18: Metabolitos identificados ordenados por classes, e respetivo desvio químico (δ)	110

Lista de abreviaturas

ANOVA	Análise de variância
AQ	Tempo de aquisição
BCG	Bacilo do <i>Calmette-Guérin</i>
BMRB	Banco de dados biológicos de ressonância magnética (Biological Magnetic Resonance Bank)
C	Grupo experimental de indivíduos saudáveis (grupo controlo)
CP 1/ 2	1 ^a componente principal/ 2 ^a componente principal
CPMG	Carr-Purcell-Meiboom-Gill
CV	Validação cruzada
D	Grupo experimental de indivíduos com tuberculose
DGS	Direção-Geral de Saúde
DEP	Grupo experimental de indivíduos com tuberculose extrapulmonar
DP	Grupo experimental de indivíduos com tuberculose pulmonar
DNA	ácido desoxirribonucleico
DSS	ácido 4,4-dimetil-4-silapentano-o-sulfónico

EDTA	ácido etilenodiaminotetracético,
ex	Exemplo
FDR	Taxa de descoberta de falsos positivos
FID	Decaimento de indução livre
<i>g</i>	Unidade de campo gravitacional, utilizada como unidade de força centrífuga relativa
G+C	Guanina mais citosina
HCl	ácido clorídrico
HIV	Vírus da imunodeficiência humana
HMDB	base de dados do metaboloma humano (Human Metabolome Database)
Hz	Hertz
IFN	Interferão
IGRA	teste de libertação do interferão gama (Interferon Gamma Release Assay)
IL	Interleucina
<i>J</i>	Constante de acoplamento (Hz)
K	Grau Kelvin

kDa	kiloDalton
LAM	Lipoarabinomananos
LB	Ampliação da linha do espectro
MDR-TB	Estirpes de <i>Mycobacterium tuberculosis</i> multirresistentes a antibióticos
MHC I/II	Complexo principal de histocompatibilidade da classe I/II
MS	Espectrometria de massa
MSEA	Análise de enriquecimento do conjunto de metabolitos
Mtb	<i>Mycobacterium tuberculosis</i>
NMR	Ressonância magnética nuclear
NaCl	Cloreto de sódio
NS	Número de scans
NOESY	Espectroscopia do efeito nuclear de Overhauser
NPV	Valor preditivo negativo
OMS	Organização Mundial de Saúde
PAMPS	Padrões moleculares característicos dos microrganismos

PCA	Análise de componentes principais
PCR	Reação em cadeia da polimerase
pH	Potencial hidrogeniónico
PIM	Manosídeos de fosfatidilinositol
PLS-DA	Análise discriminante do princípio dos mínimos quadrados parciais
ppm	Parte por milhão
PPV	Valor preditivo positivo
RD	Tempo de relaxação
RNA	Ácido ribonucleico
SMDB	Base de dados do metaboloma do soro sanguíneo humano (Serum Metabolome Database)
SW	Largura do espectro
T₁	Relaxação longitudinal
T₂	Relaxação transversal
TB	Tuberculose
TD	Tamanho do campo
TMS	Tetrametilsilano

TOCSY	Espectroscopia de correlação total
TSP	ácido trimetilsililpropanoico
TST	Teste cutâneo da tuberculina
UV	Ultravioleta
VIP	Variáveis importantes na projeção
vs	Versus
XDR-TB	Estirpes de <i>Mycobacterium tuberculosis</i> extensivamente resistentes a antibióticos
mL	Mililitro
mm	Milimetro
mM	Milimolar
μl	Microlitro
μm	Micrómetro
°C	Grau celsius

Capítulo I: Introdução

1.1 Tuberculose, uma doença emergente

Segundo a Organização Mundial de Saúde (OMS), a tuberculose continua a ser uma das principais causas de morte, afetando cerca de 1/3 da população mundial.

A tuberculose, é uma doença bacteriana que tem como agente etiológico *Mycobacterium tuberculosis* (*Mtb*). Esta doença, apesar de muito antiga, ainda causa muitos problemas de saúde pública a nível mundial. A tuberculose tornou-se uma doença emergente devido a variados fatores, como a degradação das condições sanitárias nas grandes cidades e o aumento do fluxo de migrantes. Além destes fatores acresce o fato de terem surgido estirpes resistentes aos antibióticos usados para o combate da tuberculose, a coinfeção com o vírus da imunodeficiência humana (HIV) e o abandono dos programas de controlo da doença em alguns países (Barroso et al. 2014). O aumento dos casos de tuberculose está também associado ao aparecimento de estirpes multirresistentes (MDR-TB) e de estirpes extensivamente resistentes (XDR-TB) (Barroso et al. 2014).

Segundo a Direção-Geral da Saúde (DGS), foram detetados, em 2014, cerca de vinte novos casos por cada cem mil habitantes (Diniz et al. 2015), sendo este valor baixo quando comparado com os valores relativos a países como a África do Sul e a Namíbia (Figura 1).

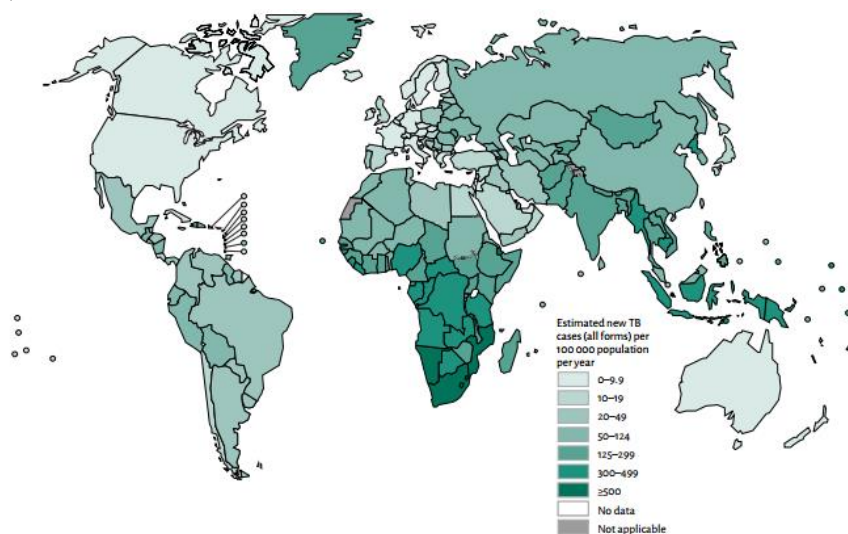


Figura 1: Taxa de incidência estimada da tuberculose em 2014 (Lewandowski et al. 2015)

De 2004 a 2014, tanto a taxa de incidência como a taxa de notificação dos casos de tuberculose, diminuíram para metade (**Erro! Autorreferência de marcador inválida.**). As regiões de Lisboa, Setúbal, Algarve e Porto, registaram as maiores taxas de incidência, neste período (Diniz et al. 2015).

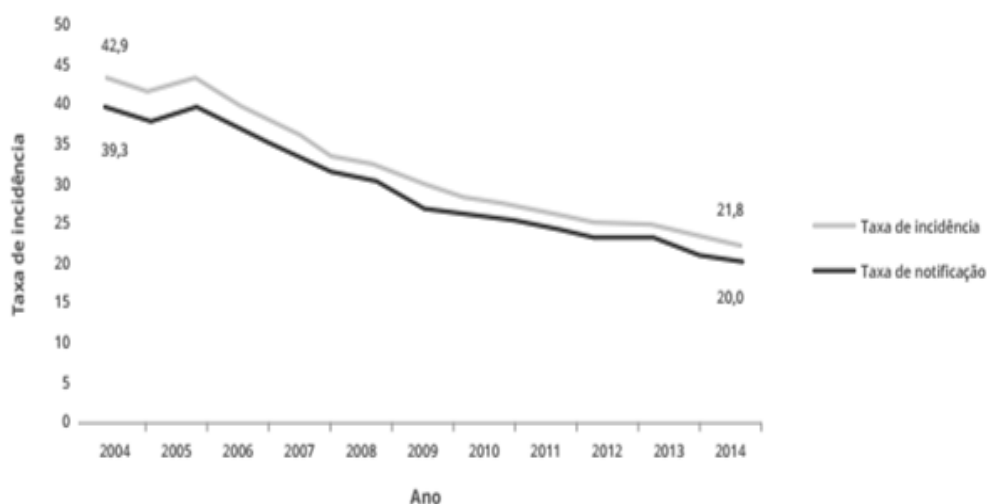


Figura 2: Evolução da taxa de notificação e de incidência de tuberculose em Portugal, 2004-2014 (Diniz et al. 2015)

1.2 O género *Mycobacterium*

As bactérias pertencentes a este género fazem parte da família *Mycobacteriaceae*, que está incluída na ordem *Actinomycetales*. Este género é filogeneticamente próximo dos géneros *Nocardia*, *Rhodococcus* e *Corynebacterium* (Barroso et al. 2014).

Este género engloba bacilos aeróbios estritos, imóveis e que tendem a agrupar-se em pequenas cordas ou aglomerados. Estas bactérias apresentam coloração Gram variável, sendo classificadas como bactérias álcool-ácido resistentes (BAAR). As bactérias pertencentes a este género também têm a particularidade de apresentar ácidos micólicos nas suas paredes celulares (Barroso et al. 2014). Apresentam também elevado teor (61 – 71mol%) de guanina mais citosina (G+C) no DNA (Murray et al. 2011).

Dentro deste gênero, encontram-se descritas cerca de cento e vinte e seis espécies (Barroso et al. 2014).

O gênero *Mycobacterium* encontra-se dividido em dois grandes grupos. Um grupo compreende bactérias de crescimento lento e o outro grupo inclui bactérias de crescimento rápido. As micobactérias também podem ser classificadas segundo o risco de infecção e de biossegurança. *M. tuberculosis* e *Mycobacterium leprae*, por exemplo, são bactérias estritamente patogênicas, de crescimento lento e não pigmentadas (Barroso et al. 2014).

1.2.1 *Mycobacterium tuberculosis*

A bactéria *M. tuberculosis* foi identificada pela primeira vez por Robert Koch em 1882 (Lawn & Zumla 2011). *M. tuberculosis* é um bacilo imóvel, não esporulado e com extremidades arredondadas. Esta bactéria cresce a 37°C, sendo o Homem o seu único hospedeiro natural. As suas colónias, em meio sólido, são de cor esbranquiçada em forma de couve-flor (JPC et al. 2000).

Mycobacterium tuberculosis apresenta uma parede celular rica em lípidos e ácidos micólicos. Além destes dois componentes a parede celular é constituída também por glicolípidos ricos em ácidos micólicos, peptidoglicanos, lipoarabinomananos (LAM), manósídeos de fosfatidilinositol (PIM), dimicocerato de fitiocerol, fator corda, sulfolípidos e cera-D. Estes componentes permitem a sobrevivência da bactéria dentro dos fagossomas (Meena & Meena 2011) (Figura 3).

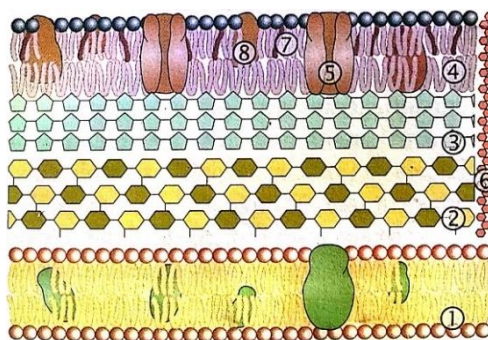


Figura 3: Representação da parede celular de Mtb. 1. Bicamada fosfolipídica com proteínas integradas; 2. Peptidoglicanos; 3. Arabinogalactano; 4. Ácidos micólicos; 5. Porinas externas; 6. Lipoarabinomananos (LAM); 7. Lípidos e glicolípidos; 8. Proteínas de superfície (Barroso et al. 2014).

Este bacilo apresenta uma replicação lenta de cerca de quinze a vinte horas. Esta característica e o facto de a micobactéria persistir, num estado de latência, no organismo humano, torna a sua eliminação uma tarefa difícil.

São vários os fatores que conferem a esta micobactéria uma virulência extrema. Os sistemas de secreção de proteínas ESX-1 e as proteínas secretadas do complexo antigénico 85, são dois dos fatores responsáveis pela virulência da bactéria. Estes sistemas de secreção permitem que a micobactéria consiga sobreviver dentro do macrófago infetado (Meena & Meena 2011). Por outro lado, as proteínas do complexo antigénico 85, são responsáveis pela biossíntese da parede celular das micobactérias (Kremer et al. 2002).

Quando em contacto com o hospedeiro, a micobactéria utiliza os granulomas formados durante a infeção, que normalmente têm a função de impedir a proliferação da micobactéria, como uma fonte rica em nutrientes para *Mtb*. A bactéria consegue recrutar novas células, que ainda não foram infetadas, para o local da infeção, através do locus bacteriano RD1 (Meena & Meena 2011).

1.3 Patogénese e Imunologia

A transmissão da tuberculose, por via respiratória, ocorre quando indivíduos com tuberculose pulmonar ativa, expelem partículas com bacilos de *M. tuberculosis* através da tosse e/ou espirros (Delogu et al. 2013; Fogel 2015). Estas partículas designadas de gotículas de *Pflugge* (Lesch 1996), por ação do vento, do calor e da radiação UV, vão desidratar, dando origem a partículas de menores dimensões (1 a 2 µm) (Barroso et al. 2014) denominadas de núcleos de *Wells*. Estes núcleos têm a particularidade de serem extremamente pequenos, conseguindo ultrapassar todas as barreiras do trato respiratório, atingindo os alvéolos pulmonares. A transmissão da tuberculose vai ser determinada pela infecciosidade do doente, pela concentração do microrganismo presente nas gotículas de *Pflugge*, pelo tempo de exposição ao microrganismo e pelo estado imunológico dos

indivíduos que se encontram na proximidade dos doentes (contactos) (Barroso et al. 2014).

O bacilo ao chegar aos alvéolos pulmonares é de imediato fagocitado pelos macrófagos alveolares e pelas células dendríticas (Fogel 2015). O contacto do macrófago com a micobactéria ocorre através dos recetores de reconhecimento de padrões, como é o caso dos recetores *toll-like* que reconhecem padrões moleculares característicos dos microrganismos (PAMPS), como os LAMs expressos por *Mtb*. Após este contacto, o macrófago irá fagocitar o bacilo (Fogel 2015). Ao entrar dentro do macrófago, a bactéria encontra-se inicialmente numa vesícula denominada de fagossoma que, ao fundir-se com os lisossomas, irá originar fagolisossomas que contêm intermediários reativos de oxigénio, enzimas lisossomais e péptidos tóxicos, apresentam um pH baixo, que tornam o fagolisossoma um agente microbicida (Smith 2003). Após destruir a micobactéria, o macrófago irá expressar moléculas do tipo MHC I, responsáveis pela apresentação de pequenos péptidos do bacilo às células T CD8+. Aquando do contacto dos recetores *toll-like* com os PAMPS da micobactéria, os macrófagos irão produzir citocinas, como a IL-1, a IL-6 e a TNF- α que vão promover a vasodilatação, o aumento da permeabilidade vascular, a quimiotaxia dos neutrófilos e a diapedese (transmigração das células do sangue para o tecido) (Kaufmann & Parida 2008).

As células dendríticas, após fagocitose, expressam moléculas do tipo MHC II e vão apresentar os péptidos do bacilo às células T CD4+ (Kaufmann & Parida 2008). As células dendríticas, vão ter aos nódulos linfáticos, onde se encontram as células T (Fogel 2015). Ao serem ativadas, as células T podem diferenciar-se em células Th1 e Th2. As células Th1, diferenciadas através da IL-12 produzida pelas células dendríticas, produzem IFN- γ para estimular os macrófagos a exercer a sua função microbicida. Por sua vez, os macrófagos produzem IL-12 para agir sobre as células T CD4+, criando um ciclo (Kaufmann & Parida 2008).

Após a ativação da resposta imune adaptativa, os macrófagos passam a produzir uma maior quantidade de citocinas. Contudo, as micobactérias conseguem sobreviver dentro dos macrófagos através do desencadeamento de uma resposta inflamatória, do bloqueamento da produção de intermediários reativos do oxigénio e do nitrogénio e, ainda, através da redução da acidificação do fagossoma (Sasindran & Torrelles 2011). Assim a bactéria multiplica-se dentro do macrófago, induzindo a sua morte e

disseminando-se. Monócitos e outras células inflamatórias serão para o local da infecção (Sasindran & Torrelles 2011). Após a contínua produção de citocinas, as células T criam uma barreira em volta dos macrófagos, limitando a sua mobilidade. Assim irá formar-se um granuloma que irá posteriormente ser composto por macrófagos epiteliais e por material fibrótico que irá calcificar (Fogel 2015).

Em granulomas pequenos, as bactérias intracelulares são eliminadas. Já no caso de granulomas grandes, estes calcificam, não permitindo a eliminação da bactéria. Nos granulomas maiores, as micobactérias tanto podem permanecer num estado de dormência durante toda a vida do hospedeiro, como podem ser reativadas posteriormente. A reativação da infecção vai ser causada por distúrbios no sistema imune do hospedeiro (Murray et al. 2011).

A ocorrência da disseminação hematogénica das micobactérias durante o estágio inicial da infecção, origina casos de tuberculose extrapulmonar (Murray et al. 2011).

1.4 Diagnóstico

O diagnóstico da tuberculose envolve quatro etapas sucessivas que passam pela deteção da infecção latente da tuberculose, pela triagem da tuberculose ativa, pela confirmação da tuberculose ativa e pela determinação da espécie do organismo infetante (du Preez & Loots 2014).

O teste *Mantoux* ou TST e o teste da libertação do interferão gama (IGRA) são os métodos de diagnósticos mais utilizados na deteção da tuberculose. O teste *Mantoux* visa a medição da resposta imune do organismo 48 a 72 horas após a injeção subcutânea de derivados da tuberculina. Uma reação positiva apresenta um eritema e uma induração superior a 10 mm (du Preez & Loots 2014; Murray et al. 2011). O teste do IGRA é amplamente utilizado para a deteção da tuberculose latente. Este método de diagnóstico verifica a produção de interferão gama produzidos pelas células T. Estas células vão ser estimuladas a produzir interferão gama através da utilização de antígenos da *Mtb*, como é o caso do antígeno ESAT-6 (antígeno alvo secretado precocemente), do CFP-10

(proteína 10 filtrada em cultura) e do TB7.7. Os antígenos ESAT-6 e CFP-10 são codificados na região RD1 do genoma e o antígeno TB7.7 é codificado numa inserção do genoma de *Mtb* (Pai et al. 2004)

O teste *Mantoux* apesar de ser económico, não é conclusivo quando aplicado a indivíduos que tenham sido vacinados com a BCG. Este facto, deve-se à reação cruzada com o bacilo de *Calmette-Guérin* e outras bactérias não tuberculosas. (Nayak & Acharjya 2012). Os testes de imunodiagnóstico, como o teste de *Mantoux* e o teste do IGRA, apresentam algumas limitações. Estas limitações passam pelos resultados negativos gerados em indivíduos imunocomprometidos. Além disso, um resultado positivo no teste do IGRA não indica obrigatoriamente um diagnóstico positivo de TB ativa, sendo necessário a realização de testes de diagnósticos adicionais para a sua confirmação (du Preez & Loots 2014).

Outros testes que se podem realizar para o diagnóstico de TB estão descritos na Tabela 1.

Tabela 1: Métodos de diagnósticos usados correntemente. PPV: valores preditivos positivos, NPV: valores preditivos negativos, MODS: suscetibilidade a drogas observada por microscopia, E-MTD: potenciador direto de *Mycobacterium tuberculosis* (du Preez & Loots 2014).

Métodos de diagnóstico	Sensibilidade (%)	Especificidade (%)	PPV (%)	NPV (%)	Limite de deteção (bactérias/mL)	Tempo de diagnóstico	Capacidade de especiação
Raio-X	91	67	78	84	---	1 hora	Não
Baciloscopia	62	99	96.4	84.2	10-100	2 horas	Não
Cultura Lowenstein-Jensen	84	100	99.7	98.1	10-100	26 dias	Sim
MBBacT	89	99	98.9	98.7	100	13 dias	Sim
MODS	97.4	94.4	92.6	98.1	---	7 dias	Sim
E-MTD	83	97	88	95	2	3 horas	Não
Taqman MTB PCR	79.1	98.2	73.1	98.7	18	2 ½ horas	Não
Amplicor MTB PCR	58.3	99.5	87.5	97.4	100	6 ½ horas	Não
Ensaio FAST-Plaque Phage	72.5	99	91	96	---	2 dias	Não
MeDiPro <i>M. tuberculosis</i>	81.8	96.7	98.3	69.5	150µl de fluido de cultura positiva	20 minutos	Não
Teste rápido de antígeno							

Os testes de cultura de *Mtb* e de identificação de resistências a antibióticos têm a agravante de serem técnicas demoradas, sendo apenas possível obter resultados após 4 a 6 semanas, em meio de cultura sólido, e 2 a 4 semanas, em meio de cultura líquido. Os sistemas de cultura automatizados, como o BACTEC MGIT 960 e MBacT apesar de gerarem resultados com maior rapidez, têm a desvantagem de serem demasiado dispendiosos. O teste MODS, por seu lado, tem ainda a desvantagem de necessitar de observações microscópicas diárias (du Preez & Loots 2014).

A deteção da infeção por tecnologias de amplificação de ácidos nucleicos, apresenta algumas desvantagens. Estas desvantagens passam pelo facto de este método gerar falsos negativos devido à presença de inibidores do PCR nas amostras biológicas utilizadas. Além disso estas técnicas geram também falsos positivos causados pela presença de vestígios de *Mtb* deixados por uma infeção anterior (du Preez & Loots 2014).

Outros testes utilizados no diagnóstico da tuberculose, apresentam uma baixa sensibilidade ou uma baixa reprodutibilidade, não sendo recomendada a sua utilização (du Preez & Loots 2014).

Muitos dos diagnósticos descritos na Tabela 1, não podem ser utilizados individualmente, sendo necessário serem complementados com outros métodos de diagnóstico. Um método de diagnóstico ideal precisa ser rápido, simples e de elevada sensibilidade. Além disso, necessita de ter uma especificidade e valores preditivos positivos (PPV) de aproximadamente de 100%. Precisa também de ser capaz de identificar e conseguir diferenciar as várias espécies de *Mycobacterium* e de conseguir caracterizar a sua suscetibilidade a antibióticos (du Preez & Loots 2014).

Dada as limitações dos métodos de diagnóstico utilizados hoje em dia, é necessário continuar a desenvolver novas ferramentas para detetar a tuberculose. Têm vindo a ser desenvolvidos estudos de metabolómica para a criação de um método de diagnóstico ótimo. Alterações causadas, por exemplo pela infeção de *Mtb*, no organismo humano, vai refletir-se no seu metaboloma. Isto porque, alterações que ocorram ao nível do genoma, do transcriptoma e do proteoma humano, vão potencialmente afetar o metaboloma (du Preez & Loots 2014)

1.5 Tratamento

O tratamento da tuberculose é realizado atualmente através da administração de 5 antibacilares: isoniazida, rifampicina, estreptomicina, etambutol e pirazinamida. Nos primeiros 2 meses são administrados a isoniazida, a rifampicina, a pirazinamida e o etambutol (Figura 4). Nos 4 meses seguintes apenas são administrados a isoniazida e a rifampicina. A estreptomicina apenas é administrada em caso de meningite tuberculose, substituindo o etambutol. Estes fármacos pertencem à primeira linha de medicamentos. Caso estes fármacos falhem opta-se pela administração de fármacos de 2ª linha que levam a reações severas. Os fármacos de 2ª linha apenas são administrados a indivíduos com estirpes multirresistentes (Barroso et al. 2014). Em Portugal, como se pode observar pela tabela 2 ocorreram alguns casos de resistência aos medicamentos de primeira linha, no ano de 2014.

Tabela 2: Número de casos com resistência aos antibacilares de primeira linha em Portugal em 2014 (Diniz et al. 2015).

Isoniazida	Rifampicina	Etambutol	Estreptomicina	Isoniazida + Rifampicina
97	25	15	124	23

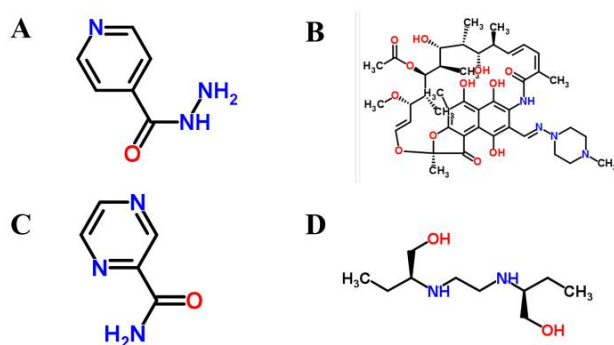


Figura 4: Estrutura química dos antibacilares de primeira linha utilizados no tratamento da tuberculose. A: Isoniazida; B: Rifampicina; C: Pirazinamida; D: Etambutol

1.6 O sangue, um biofluido de excelência

1.6.1 Constituição sanguínea

O sangue é um fluido biológico complexo constituído por plasma e componentes celulares (hemácias e leucócitos). Este fluido é responsável pelo transporte de várias componentes como, iões minerais, proteínas, hormonas, plaquetas, dióxido de carbono, lípidos e outros compostos de baixa massa molecular. Cerca de 91% do sangue é apenas água. Apesar de apresentar uma grande variedade de constituintes, o sangue só é responsável pelo seu transporte, pela regulação do pH e da temperatura corporal (Guthrie 2012).

1.6.2 Plasma versus soro

Em estudos de metabolômica não é recomendada a utilização do sangue por inteiro. Isto porque este biofluido apresenta uma fraca estabilidade, é heterogéneo e não é compatível com a maioria dos métodos usados para a preparação de amostras (Vuckovic 2013). Em vez disso é utilizada a parte líquida do sangue correspondente ao plasma ou ao soro. Além disso, o fracionamento do sangue permite a simplificação da amostra, dado que são removidas as células sanguíneas e algumas proteínas.

O plasma, correspondente à fração líquida do sangue, é obtido através da adição de anticoagulantes, como a heparina ou o EDTA, para que não ocorra a formação do coágulo sanguíneo. O soro, tal como o plasma, também corresponde à fração líquida do sangue. A diferença entre o soro e o plasma, é que para a obtenção do primeiro deixa-se ocorrer o processo de coagulação do sangue. Assim, as proteínas envolvidas no processo de coagulação não vão fazer parte da constituição do soro. Quer para o plasma, quer para o soro é realizado um passo de centrifugação para separar a componente celular (Vuckovic 2012).

Existem uma grande variedade de diferenças entre o soro e o plasma descritas na literatura, tais como:

- Níveis aumentados de lisofosfatidilinositol no plasma;
- Tromboxano B2 presente no soro e ausente no plasma (Wedge et al. 2011);
- Concentração aumentada de aminoácidos no soro, como é o caso da arginina, da serina, da fenilalanina e da glicina (Brauer et al. 2011);
- Níveis aumentados de certas fosfatidilcolinas no soro (Yu et al. 2011);
- Presença de fibrinogénio no plasma, estando ausente no soro.

Para a identificação e quantificação de metabolitos em estudos de metabolómica, o soro é uma das amostras biológicas mais utilizadas (Guthrie 2012). Para além de mais, o soro é fácil de recolher, quando comparado com o plasma, dado que na recolha do soro não é necessário recorrer a uma centrifugação refrigerada (Vuckovic 2012).

1.6.3 Metaboloma do soro

Atualmente existem diversos estudos publicados, onde foi realizada a caracterização do metaboloma do soro sanguíneo. Muitos destes estudos utilizam técnicas como a espectrometria de massa (MS) (Psychogios, David D Hau, et al. 2011; Boudah et al. 2014; Dunn et al. 2011) e a ressonância magnética nuclear (NMR) (Mickiewicz et al. 2014; Yu et al. 2011; Zhou et al. 2013), dois dos métodos mais utilizados na área da metabolómica.

A partir da base de dados do metaboloma do soro sanguíneo humano (SMDB:// www.serummetabolome.ca) é possível encontrar uma lista de metabolitos descritos em vários estudos publicados na área de metabolómica (Psychogios, David D. Hau, et al. 2011). Até à data a SMDB contém informações sobre 4549 metabolitos detetáveis no soro.

Foram também realizados estudos metabolómicos para verificar alterações na composição/concentração do metaboloma do soro sanguíneo causadas pela idade, género,

índice de massa corporal e hábitos tabágicos (Mittelstrass et al. 2011; Dunn et al. 2014; Yu et al. 2012; Moore et al. 2014).

1.7 Metabolômica

Nas últimas décadas tem-se verificado um aumento do número de estudos realizados com base nas tecnologias das “ômicas”, como é o caso da genômica, da transcriptômica, da proteômica e da metabolômica (Figura 5). A genômica, a transcriptômica e a proteômica têm como objetivo o estudo de todos os genes, transcritos e proteínas, respetivamente, de uma célula, tecido, fluido, órgão ou organismo num dado momento e condição. Por seu lado, a metabolômica visa estudar o perfil metabólico de uma célula, tecido, fluido, órgão ou organismo num dado momento e condição. A lipidômica, a glicômica e a fluxômica são mais alguns exemplos de “ômicas” (Ma et al. 2012; Leibfritz et al. 2007; Nielsen & Oliver 2005).

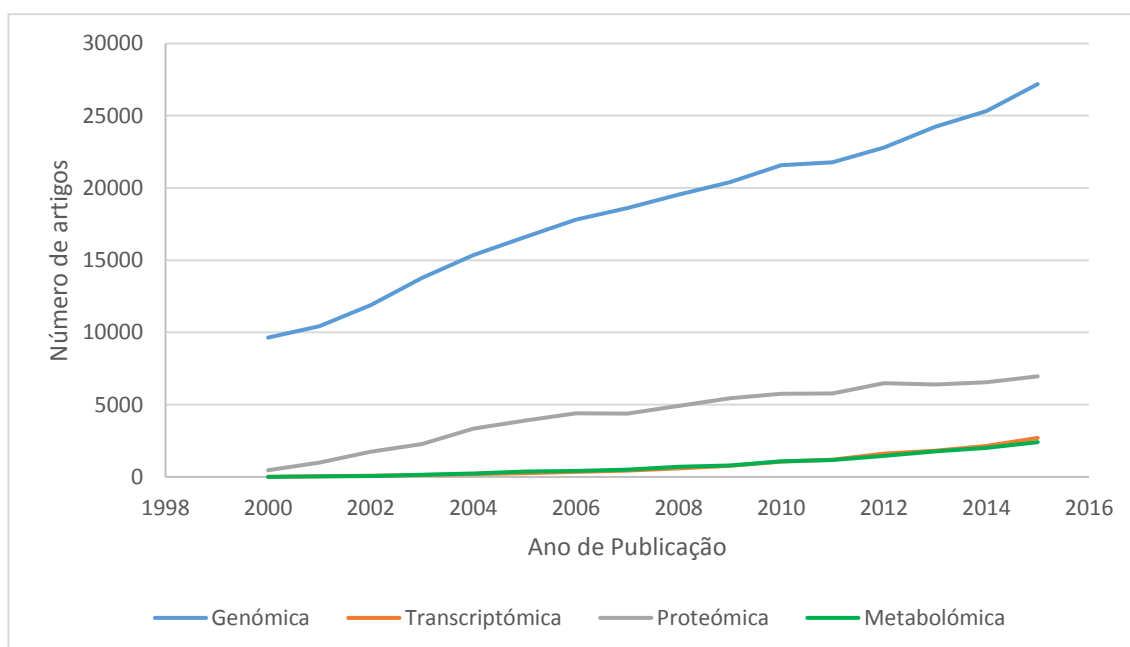


Figura 5: Gráfico representativo do número de artigos publicados com os respetivos termos ("genomics", "transcriptomics", "proteomics" e "metabolomics") desde 2000 até 2015. A pesquisa do número de publicações em cada área foi realizada através do site ScienceDirect (www.sciencedirect.com)

A metabolômica oferece várias vantagens, quando comparada com as restantes ômicas, sendo a abordagem que mais se aproxima do fenótipo de um sistema biológico, permite a rápida observação de potenciais perturbações ocorridas no metaboloma (Zhang et al. 2012).

Segundo Oliver e colaboradores a metabolômica define-se como sendo o estudo do “conjunto completo de metabolitos/ intermediários de baixa massa molecular, os quais são dependentes do contexto, variando de acordo com o estado fisiológico, de desenvolvimento ou patológico da célula, tecido, órgão ou organismo” (Oliver 2003). Moléculas como glícidos, vitaminas, lípidos, polifenóis, aminoácidos e péptidos são considerados metabolitos. O conjunto de metabolitos de uma célula, órgão ou organismo é definido como metaboloma (Wishart 2008).

A metabolômica engloba 2 abordagens, a “*targeted*” e a “*untargeted*”. A abordagem “*targeted*” visa a deteção e quantificação específica de metabolitos que participam numa ou mais vias metabólicas alvo. Dentro da abordagem “*untargeted*” existem dois tipos de análise: o perfil metabólico e o “*fingerprinting*” de metabolitos. O perfil metabólico visa a deteção, quantificação absoluta e identificação de metabolitos presentes numa amostra biológica. Por outro lado, o “*fingerprinting*” de metabolitos apenas visa a deteção comparativa e quantificação relativa de metabolitos em amostras de dois grupos distintos (ex: controlos vs doentes) (Villas-Boas et al. 2006).

A metabolômica, no campo da medicina, tem vindo a ser aplicada para a descoberta de novos biomarcadores, para a quantificação da resposta terapêutica e para a caracterização do perfil metabólico de doenças (Kaddurah-Daouk et al. 2008). A metabolômica tem vindo a ser aplicada na medicina dado que através desta área é possível fazer uma identificação não dirigida de várias dezenas a centenas de metabolitos de uma única vez. Além disso, a metabolômica permite traçar o perfil de doenças, possibilitando a compreensão das mesmas (Banoei et al. 2014).

Metodologias como a espectrometria de massa (MS) e a ressonância magnética nuclear (NMR), têm visto a sua aplicação aumentada na área da metabolômica. São utilizadas desde os anos 70 para estudos de doenças e de mecanismos biológicos (Devaux et al. 1971; Fan 1996).

1.8 Ressonância magnética nuclear (NMR)

O spin é uma propriedade quântica que, dependente da composição de cada núcleo, apresenta valores de 0, $\frac{1}{2}$ ou superiores. Os núcleos atômicos com spins de $\frac{1}{2}$ são idealmente utilizados para a ressonância magnética nuclear (NMR), apresentando propriedades ideais para a sua detecção. Os núcleos ^1H , ^{13}C , ^{15}N e ^{31}P , presentes em sistemas biológicos, apresentam spins de $\frac{1}{2}$. Estes núcleos apresentam spins diferentes de zero, funcionando como barras magnéticas.

Na natureza, os spins encontram-se orientados de forma aleatória e têm igual energia. Mas, quando aplicamos um campo magnético sobre os núcleos, estes vão apresentar spins com diferentes níveis de energia. Para núcleos com spins de $\frac{1}{2}$, estes irão ser divididos em dois níveis de energia ($-\frac{1}{2}$ e $+\frac{1}{2}$). A diferença de energia entre os dois níveis é muito reduzida quando comparada com o nível de energia ambiental. Como consequência, os spins distribuem-se igualmente entre os dois níveis de energia (Transfer & Restraint 1970).

Quando aplicamos um campo magnético sobre os núcleos, estes vão apresentar spins com diferentes níveis de energia. Se direcionarmos ondas eletromagnéticas na região da radiofrequência, os núcleos com spins com níveis de energia mais baixo ($-\frac{1}{2}$) ganham energia suficiente para transitar para níveis de energia mais altos ($+\frac{1}{2}$). Quando voltam ao estado de equilíbrio ($-\frac{1}{2}$), os núcleos excitados libertam radiação na gama das radiofrequências que é detetada pelo espectrómetro de ressonância magnética nuclear. A esta frequência da radiação emitida pelo núcleo é designada de frequência ressonante ou desvio químico (Transfer & Restraint 1970). Os spins ao voltarem ao seu estado de equilíbrio, vão emitir energia que vai ser detetada sob a forma de um sinal de decaimento de indução livre (FID), no domínio do tempo. O FID vai dar origem a um espectro através da transformação de Fourier (Rolin et al. 2013).

Através da análise dos parâmetros medidos no espectro de NMR é possível obter informações sobre a estrutura molecular da amostra. O desvio químico (δ em partes por

milhão, ppm), a multiplicidade de sinais, a constante de acoplamento (J) e a intensidade de sinais são os parâmetros que podem ser medidos nos espectros de NMR para se obter informação estrutural (Fan 1996). O desvio químico de qualquer núcleo atômico vai depender do seu ambiente vizinho e da sua nuvem eletrônica, indicando o grupo químico ao qual pertence a molécula. A multiplicidade de sinais vai depender do número de prótons acoplados ao núcleo observado. A constante de acoplamento é uma constante que vai depender das ligações eletrônicas com núcleos vizinhos. A intensidade dos sinais vai depender do número de prótons na amostra. Quanto mais concentrado estiver um composto na amostra, maior irá ser a intensidade do seu sinal (Fan 1996).

O tempo de relaxação longitudinal (T_1) e o tempo de relaxação transversal (T_2) são outros dois parâmetros que podem ser medidos no espectro de NMR. Moléculas de menor massa molecular iram apresentar um T_1 maior, comparativamente com moléculas de grande massa molecular. O tempo de relaxação vai influenciar a intensidade e a forma dos sinais (Deborde et al. 2013).

O campo magnético vai ser criado por magnetes selenoides supercondutores. Os selenoides estão imersos em hélio líquido a 4K e estão cercados por azoto líquido a 77K, que por sua vez está rodeado por uma câmara de vácuo. O magnete é atravessado por um tubo à temperatura ambiente onde se encontram as bobinas elétricas que geram um campo magnético próprio, homogeneizando o campo magnético (Gil & Geraldés 2002).

O computador vai servir para controlar as sequências de pulso e para fazer a transformação de Fourier. Além disso, é possível armazenar e processar os dados de NMR no computador. A consola vai transmitir os pulsos de radiofrequência, recebendo e amplificando o sinal de resposta, posteriormente. O magnete supercondutor providência um campo magnético, forte e homogêneo no qual a amostra irá ser colocada. A sonda, por sua vez, funciona como uma antena, que permite a transmissão da radiofrequência entre a amostra e os instrumentos eletrônicos (Figura 6).

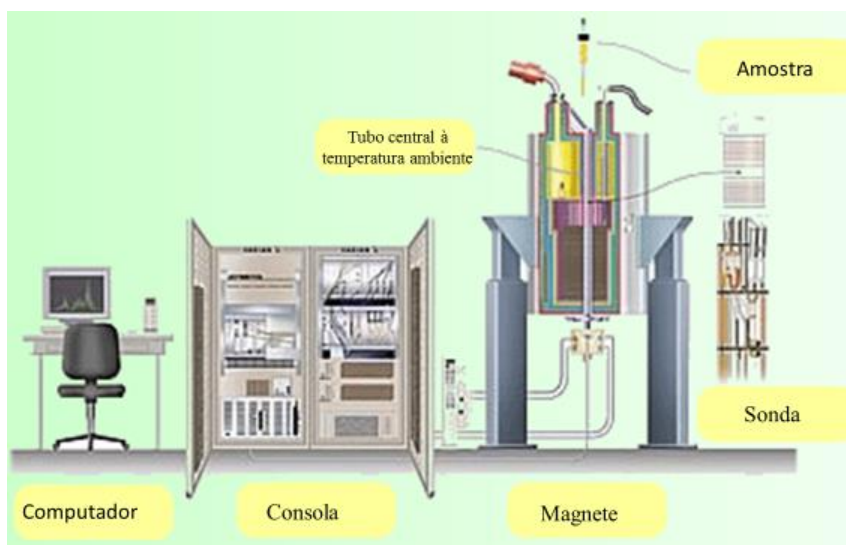


Figura 6: Componentes de um espectrómetro de ressonância magnética nuclear

Os espectros de uma dimensão (1D) são muito utilizados em estudos de metabolômica. Estes espectros apresentam curtos períodos de aquisição, permitindo a análise um grande número de amostras num curto espaço de tempo. Em amostras biológicas neste tipo de experiências, é sempre necessário a realização de um passo de supressão do solvente. Dado a elevada quantidade de água na amostra biológica, o sinal deste solvente acaba por ser muito intenso, tornando os sinais dos metabolitos das regiões vizinhas do espectro pouco ou quase nada visíveis (Transfer & Restraint 1970).

As experiências 1D mais utilizadas, recorrem a sequências de pulso 1D NOESY-presat, que fornece uma boa supressão do sinal da água mantendo, ao mesmo tempo, a linha de base plana (Transfer & Restraint 1970).

A sequência de pulso CPMG também pode ser utilizada nas experiências 1D. Esta sequência não requer a prévia remoção das macromoléculas, reduzindo a manipulação da amostra. Esta sequência de pulso atenua os sinais das macromoléculas de relaxação rápida, tornando visíveis os sinais dos metabolitos de baixa massa molecular, através da edição do tempo de relaxação T2 (Transfer & Restraint 1970).

Apesar do curto período de aquisição, a identificação de metabolitos torna-se difícil em espectros 1D, devido à elevada complexidade dos espectros, causada pela sobreposição dos sinais (Viant 2003). Para alcançar a identificação dos metabolitos utilizam-se experiências 1D, em conjunto com experiências de duas dimensões (2D). Os

espectros 2D mais utilizados para estes casos são o ^1H - ^1H TOCSY, o ^1H *J-resolved* e o ^1H – ^{13}C HSQC (Brennan, L., 2014).

1.8.1 NMR na metabolómica

A partir da década de 70, a espectroscopia de NMR, começou a ser aplicada para o estudo da composição metabólica de biofluidos, células e tecidos (Wilson et al. 1974; Nicholson et al. 1983; Nicholson et al. 1985), sendo a sua utilização cada vez mais usual na metabolómica (Figura 7).

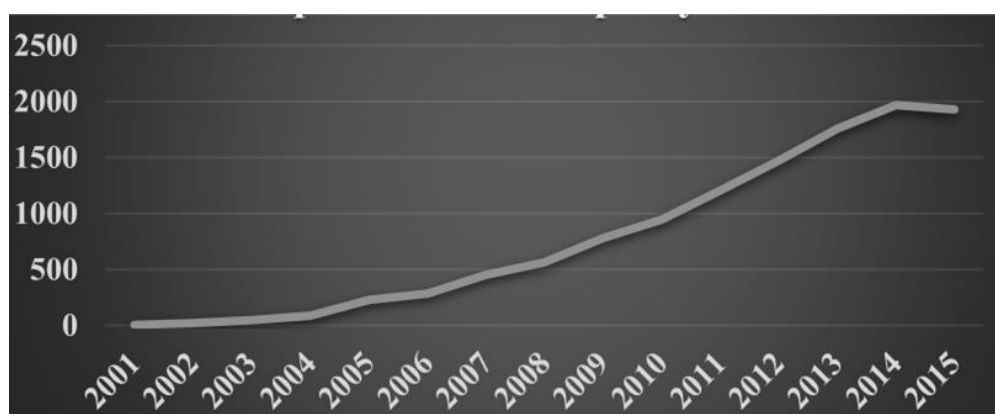


Figura 7: Gráfico representativo do numero de Publicações por ano (2001-2015) na área da metabolómica usando a técnica de NMR. Web of Science Scopus foi usado para pesquisar por este tipo de publicações usando os termos "metabolomics" e "NMR" em "All fields"

O NMR é uma técnica não destrutiva e não seletiva, permitindo replicar a obtenção de dados para uma amostra e permitindo a observação de uma grande variedade de metabolitos numa única experiência (Chadzynski et al. 2014; Sharma et al. 2015; The Economist 2014). Além disso, a sua elevada reprodutibilidade, a mínima preparação de amostra e o curto tempo de aquisição de espectros, faz com que esta técnica tenha um rendimento elevado (Winning et al. 2008).

Utilizar a espectroscopia de ressonância magnética nuclear em estudos de metabolómica como este, apresenta as suas vantagens. O NMR, é uma técnica rápida,

robusta e reprodutível (Lenz & Wilson 2007; Smolinska et al. 2012). Mas tal como a generalidade das metodologias instrumentais, o NMR também apresenta as suas desvantagens. Quando comparada com outras técnicas usadas na área da metabolómica, tais como a cromatografia líquida acoplada à espectrometria de massa (LC-MS) e à cromatografia gasosa acoplada à espectrometria de massa (GC-MS), o NMR apresenta baixa sensibilidade. Além disso apresenta maior sensibilidade para os compostos polares e, ainda, existe o facto de muitos compostos apresentarem sinais na mesma região dos espectros (Lenz & Wilson 2007; Banoei et al. 2014).

A técnica de NMR usada para estudos de metabolómica tem permitido o desenvolvimento de diversas áreas, como a medicina e a farmacologia, aplicadas no desenvolvimento de métodos de diagnóstico e na descoberta de novos medicamentos, respetivamente (Emwas et al. 2013; Powers 2009). Além disso o NMR tem sido aplicado na ciência gastronómica (Trimigno et al. 2015) e na metabolómica ambiental (Viant et al. 2003).

1.8.2 Preparação de amostras para NMR

Diferentes fluidos biológicos têm vindo a ser alvo de estudos na área da metabolómica, tais como o soro e o plasma sanguíneo (Beckonert et al. 2007; Khan et al. 2011; Diaz et al. 2011; Psychogios, David D Hau, et al. 2011; Zhang et al. 2013; Zhou et al. 2013), a urina (Beckonert et al. 2007; Khan et al. 2011; Diaz et al. 2011), a saliva e o fluido cefalorraquidiano (Levine et al. 2000; Cruz et al. 2014).

Muitos fatores, como a dieta, a administração de medicamentos e condições de saúde dos indivíduos podem afetar as concentrações metabólicas num organismo. Assim, neste tipo de estudos é deveras importante controlar as condições inerentes aos participantes no estudo, recolha e processamento das amostras (Emwas et al. 2013).

A recolha das amostras de biofluidos deve ser realizada sob condições específicas para evitar qualquer variação de dados e para evitar a presença de compostos indesejados (Smolinska et al. 2012). Para estudos como o descrito neste trabalho, é necessário ter em atenção alguns cuidados quando se lida com amostras de soro sanguíneo. O processo de

coagulação do soro deve ser realizado em gelo, durante 30 a 60 minutos, para evitar a atividade catalítica de algumas proteínas (Teahan et al. 2006). Depois da separação do soro é necessário o seu armazenamento a -80°C para interromper a atividade catalítica das proteínas presentes e contaminação da amostra por microrganismos (Issaq et al. 2009; Emwas et al. 2013).

Para que não haja alterações metabólicas durante a aquisição dos espectros de NMR é necessário remover as proteínas da amostra (Emwas et al. 2013). Existem diversos protocolos disponíveis, como a precipitação de proteínas usando solventes orgânicos (Tiziani et al. 2008; van Oss 1989; Zhang et al. 2008; Polson et al. 2003; Daykin et al. 2002) e a ultrafiltração (Tiziani et al. 2008; Daykin et al. 2002). Estudos que fizeram a comparação entre estes dois métodos descrevem a ultrafiltração como o método ideal para NMR (Daykin et al. 2002; Tiziani et al. 2008). A única desvantagem deste método reside no facto de ser necessário realizar lavagens sucessivas dos filtros, antes da filtração das amostras, para remover o glicerol presente na membrana dos filtros (Brennan 2014).

Tal como nas alterações ao nível das amostras, também é preciso ter especial cuidado com as variações das ressonâncias dos núcleos de hidrogénio dos metabolitos. Para isso é necessário ter em atenção o pH da amostra, o solvente adicionado e o composto de referência.

O pH da amostra influencia a posição dos picos dos metabolitos na região do espectro (Smolinska et al. 2012). Como o pH sanguíneo é de aproximadamente 7.7, para experiências de NMR utiliza-se um tampão fosfato a pH 7.0 evitando alterações nos desvios químicos (Beckonert et al. 2007). Para além do tampão fosfato, soluções de NaCl e de HCl também podem ser usadas (Brennan 2014).

O solvente ideal não deve apresentar sinais sobrepostos com sinais de metabolitos da amostra, devendo ser hidrofílico para conseguir dissolver a amostra, se esta for aquosa. D_2O é um dos solventes mais utilizados em metabolômica (Transfer & Restraint 1970).

Por fim deve-se ter em conta o composto de referência utilizado. Se for adicionado a uma amostra um composto de referência com concentração conhecida é possível determinar as concentrações metabólicas absolutas dos metabolitos presentes na amostra (Transfer & Restraint 1970). Além disso, este composto de referencia vai permitir corrigir diferenças instrumentais no desvio químico entre diferentes espectros (Emwas et al. 2013). O ácido 4,4-dimetil-4-silapentano-o-sulfónico (DSS), o tetrametilsilano (TMS) e

o ácido trimetilsililpropanoico (TSP) são os compostos de referência mais utilizados em estudos de metabolômica (Transfer & Restraint 1970).

1.9 Análise estatística

1.9.1 Análise univariada

Para a análise de dados de metabolômica podem ser utilizados diferentes métodos univariados. Métodos paramétricos, como o teste *t-Student* e a ANOVA, que verificam a existência de diferenças entre dois ou mais grupos, respectivamente, são muito utilizados em metabolômica. (Alonso et al. 2015; Heather et al. 2013).

Comparando dois grupos, como por exemplo doentes e controlos, com um grande número de variáveis, faz com que a análise univariada gere um grande número de falsos positivos, uma vez que não tem em conta a presença de interações entre as variáveis. Para contrariar este problema existem diversos métodos de correção, tais como o Bonferroni ou a taxa de descoberta de falsos positivos (FDR) (Alonso et al. 2015).

A caixa de bigodes é uma representação univariada usada para exibir a variação de uma variável nas amostras. Este tipo de gráficos permite a verificação dos valores máximos e mínimos de uma determinada variável, dos valores de Q1 (correspondente a 25% dos dados), Q2 (correspondente à mediana dos dados), Q3 (correspondente a 75% dos dados) e ainda permite observar a presença de *outliers*.

1.9.2 Análise multivariada

A análise multivariada ao contrário do que acontece na univariada, analisa todas as variáveis em conjunto, não permitindo a perda de informação. Além disso, gera

modelos mais estáveis e tem em conta outro tipo de variáveis que possam estar correlacionadas como o género, a idade e o índice de massa corporal.

A análise multivariada divide-se em métodos não supervisionados e supervisionados. Os métodos não supervisionados, como a análise de componentes principais (PCA), permitem observar a variação das amostras sem ter em conta qualquer informação sobre estas. Já os métodos supervisionados, como é o caso da análise discriminativa do princípio dos mínimos quadrados parciais (PLS-DA), são utilizados para identificar variáveis que estejam correlacionadas com a classe onde a amostra está inserida.

1.9.2.1 Análise de componentes principais (PCA)

A análise de componentes principais ou PCA foi introduzida por Pearson em 1901 e é amplamente aplicada em estudos de metabolómica (Ferreira et al. 1999; Nicholson et al. 2007; Smolinska et al. 2012).

Este método não supervisionado visa a transformação das variáveis originais, que estão correlacionadas umas com as outras, num conjunto menor de variáveis não correlacionadas conhecidas como componentes principais (CPs) (Sumner et al. 2003; Smolinska et al. 2012).

No PCA, grande parte da variância do conjunto de dados é explicada pelas primeiras componentes principais (Ferreira et al. 1999; Nicholson et al. 2007; Alonso et al. 2015). A primeira componente é responsável por explicar a maior variância dos dados, sendo que a segunda componente explica a segunda maior variância dos dados, e assim sucessivamente. Assim, este modelo permite reduzir a dimensionalidade dos dados sem perdas significativas de informação (Alonso et al. 2015).

O PCA é representado num gráfico de *scores* e num gráfico de *loadings*. Os *scores* são responsáveis por sumarizar os dados, exibindo a relação entre as amostras. Os *loadings* por outro lado, permitem descrever as correlações entre as variáveis originais e as componentes principais (Alonso et al. 2015).

1.9.2.2 Princípio dos mínimos quadrados (PLS-DA)

A análise discriminativa do princípio dos mínimos quadrados ou PLS-DA, é o método supervisionados mais utilizado na metabolómica, proteómica e genómica. O seu uso na metabolómica tem vindo a aumentar juntamente com o aumento dos artigos publicados nesta área (Figura 8).

Este método visa a otimização da separação entre diferentes grupos de amostras, maximizando a covariância entre as variáveis independentes X (*bins* ou concentrações metabólicas) e as variáveis dependentes Y (classes de amostras: controlos e doentes). Assim, permite-se a redução da dimensionalidade dos dados, encontrando um novo subespaço linear que vai permitir a previsão das variáveis dependentes através de um número reduzido de variáveis conhecidas como variáveis latentes (Gromski et al. 2015).

O PLS-DA permite manipular e disponibilizar dados altamente colineares. Além disso este modelo permite-nos obter gráficos de *loadings* e variáveis importantes na projeção (VIP) que permitem identificar as variáveis mais importantes responsáveis pela separação dos grupos de amostras (Gromski et al. 2015). Para identificar as variáveis responsáveis pelas maiores variações são considerados os valores de VIP superiores a 1. (Umetrics 2011).

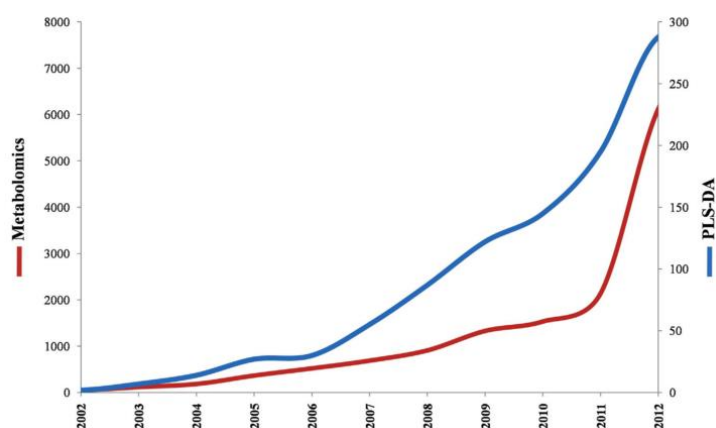


Figura 8: Gráfico representativo do número de publicações no campo da metabolómica por ano (linha vermelha) contra o número de publicações que incluem o PLS-DA como uma ferramenta na análise de dados de metabolómica (linha azul) (de Thomas Reuters' ISI Web of Science¹ usando as palavras chaves “metabolomics”, “PLS” e “PLS-DA” (Gromski et al. 2015).

1.9.2.3 Validação dos modelos estatísticos

A validação-cruzada permite estimar a capacidade de um modelo para prever corretamente a atribuição de classes para novos indivíduos. Em programas como o SIMCA P (Umetrics, Umeå, Sweden), a validação-cruzada é usada também para evitar sobre-ajustamento, através da estimativa do número de componentes que se deve usar no modelo.

A validação-cruzada envolve vários parâmetros, tais como o R^2 e o Q^2 . O R^2 indica o quão bem o modelo explica o conjunto de dados, correspondendo à fração de dados originais explicada pelo modelo. R^2 compreende valores entre 0 e 1. Caso o R^2 seja 1, indica que o modelo explica 100% dos dados. O Q^2 dá uma estimativa da capacidade preditiva do modelo. Este parâmetro é calculado pela validação cruzada *7-fold*, consistindo na divisão dos dados em sete subconjuntos. Cada um destes subconjuntos é aleatoriamente removido à vez. Deste modo pretende-se construir dois novos modelos, um com o subconjunto removido e o outro com os restantes seis subconjuntos. Isto é repetido para cada subconjunto de dados até que todos os dados tenham sido preditos. Estes irão ser comparados com os dados originais, calculando-se o erro de previsibilidade entre os modelos originais e os gerados pela validação. A soma dos erros, conhecida por PRESS é utilizada para calcular o Q^2 , sendo este parâmetro, a diferença de 1 e o valor de PRESS (Triba et al. 2015; Umetrics 2011).

Outros parâmetros como o *hotteling* T^2 e o *DModX* permitem verificar a presença de *outliers* no conjunto de dados (Broa & Smildeab 2014).

Para validar os modelos tanto de PCA como de PLS-DA, o valor de Q^2 tem de ser menor que R^2 , e a diferença entre os dois não pode ultrapassar 0,3. Valores de Q^2 e de R^2 superiores a 0,5 indicam uma boa robustez e previsibilidade do modelo. No PLS-DA é utilizado o valor de p do teste ANOVA da validação-cruzada para saber a significância estatística da separação entre os grupos (Triba et al. 2015; Umetrics 2011).

1.10. Biomarcadores no diagnóstico da tuberculose

É considerado um biomarcador qualquer tipo de substância presente num fluido ou tecido biológico que indique a presença de perturbações causadas por doença, mudanças do estado biológico ou exposição ambiental (Emwas et al. 2016).

Os biomarcadores podem ser químicos, metabolitos, genes/mutações, transcritos de RNA, proteínas, contagem de células ou tipos de células, cariótipos ou qualquer outro elemento mensurável (Emwas et al. 2016).

Um biomarcador ideal deve ser específico e seletivo para cada tipo de doença. Isto é, deve ser capaz de identificar corretamente todos os indivíduos que apresentem a doença, não gerando falsos positivos (Emwas et al. 2016).

O recurso à metabolômica na procura de biomarcadores para métodos de diagnóstico é, hoje em dia, alvo de muitos estudos. Vários estudos têm tentado detetar mudanças no metaboloma de fluidos biológicos, nomeadamente no sangue, urina, expetoração e fluido cérebroespinal, traçando um perfil metabólico da tuberculose (Haas et al. 2016). Weiner e colaboradores conseguiram identificar vinte metabolitos diferenciais entre pacientes com tuberculose e indivíduos saudáveis, com recurso à espectrometria de massa (Weiner et al. 2012). Lau e colaboradores conseguiram demonstrar que a presença do precursor de colesterol 4α -formil- 4β -metil- 5α -colesta-8-en- 3β -ol combinado com o 12-hidroxiciclopentanona ou com o sulfato de colesterol consegue diferenciar pacientes com tuberculose ativa de indivíduos saudáveis e de pacientes com pneumonia (Lau et al. 2015). Outro estudo, utilizando amostras de soro, foi realizado por Zhou e colaboradores. Neste estudo, recorreu-se ao NMR para diferenciar a composição metabólica de amostras de indivíduos com tuberculose e de indivíduos saudáveis. Através deste estudo foram detetados dezassete metabolitos em concentrações elevadas ou reduzidas em pacientes com tuberculose, indicando um consumo de energia elevado, glicólise, degradação de lípidos, biossíntese de nucleótidos e modificações do metabolismo das proteínas nos doentes (Zhou et al. 2013). Além destes estudos, outros foram publicados como o objetivo de identificar biomarcadores para o diagnóstico da tuberculose (Bagganahalli S Somashekar et al. 2012; Shin et al. 2011; Pavlou et al. 2004; Fend et al. 2006).

O problema da utilização da metabolómica para o diagnóstico, advém da heterogeneidade na variação metabólica em diferentes indivíduos e populações (du Preez & Loots 2014).

Capítulo II: Objetivos

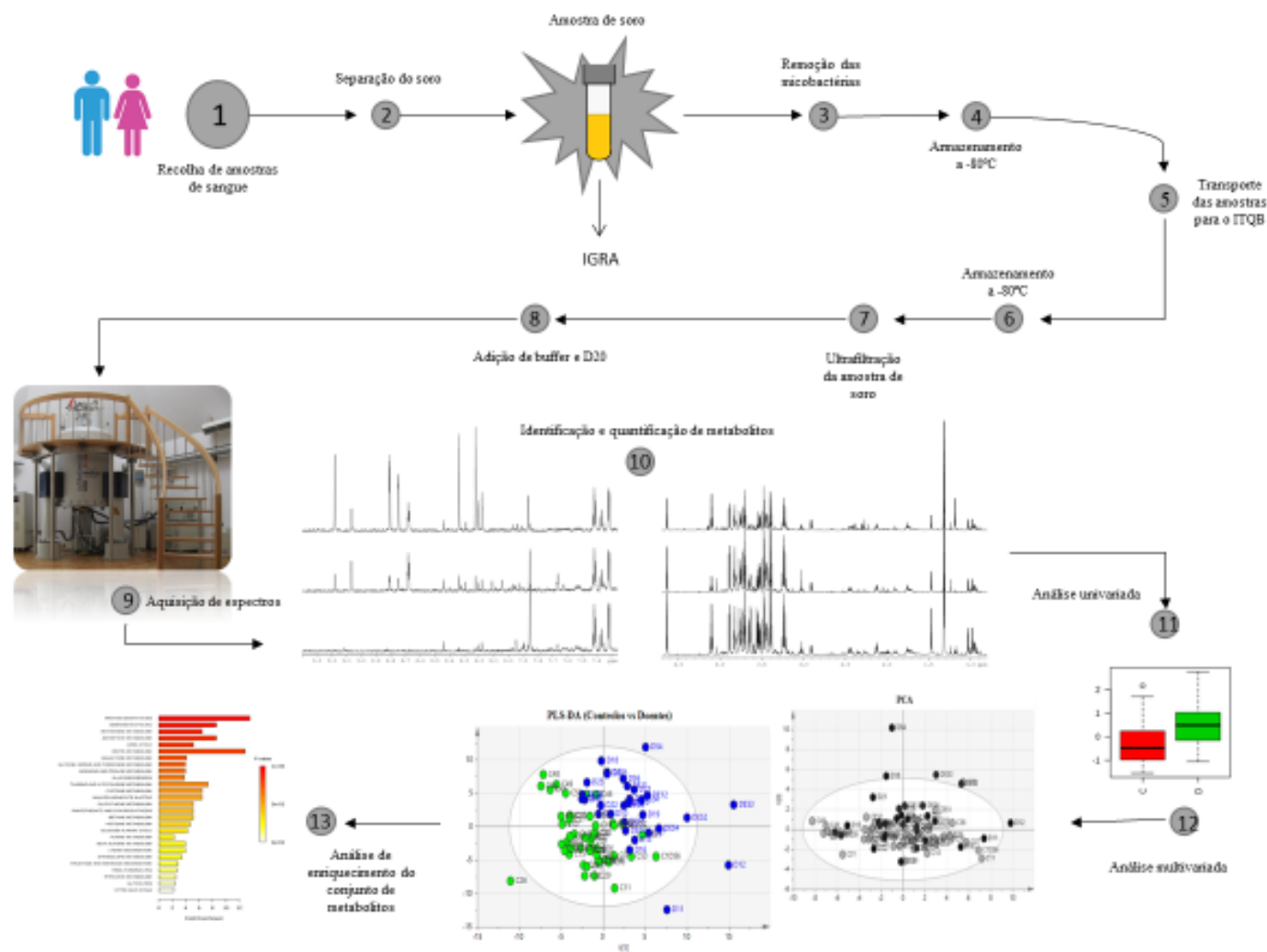
Identificar metabolitos presentes no soro sanguíneo que poderão ser utilizados como biomarcadores para o diagnóstico da tuberculose, recorrendo à ressonância magnética nuclear numa abordagem metabolómica.

- I. Otimização do procedimento para a recolha de espectros de NMR
- II. Adquirir espectros 1D e 2D de NMR de amostras de soro do grupo de doentes e do grupo controlo
- III. Aplicar um método de desconvolução para identificar e quantificar os metabolitos
- IV. Utilizar a análise *fingerprinting* para posterior identificação de metabolitos
- V. Analisar os dados através da análise univariada e multivariada
- VI. Comparar os resultados da análise *fingerprinting* e da análise do perfil metabólico
- VII. Analisar vias metabólicas envolvidas na separação do grupo de doentes e do grupo controlo

Capítulo III: Materiais e Métodos

3.1 Metodologia

Figura 9: Esquema da metodologia utilizada para este trabalho. 1: recolha de amostras de sangue; 2: separação do soro; ensaio IGRA; 3: remoção das micobactérias; 4: armazenamento a -80°C; 5: transporte das amostras para o ITQB; 6: armazenamento a -80°C; 7: ultrafiltração da amostra de soro; 8: adição de buffer e D2O; 9: aquisição de espectros; 10: identificação e quantificação de metabólitos; 11: análise univariada; 12: análise multivariada; 13: MSEA



3.2 Caracterização do perfil dos voluntários

Este trabalho decorreu no âmbito do projeto “TBomics - an OMICS approach for diagnosing tuberculosis”, New INDIGO Partnership Programme on Biotechnology applied to Human Health (INDIGO-DBT2-062), ERA-NET Project, 2014-2016 com as autorizações da Comissão de Ética da ARS de Lisboa e Vale do Tejo (Parecer 057/CES/INV/2014) e da Comissão Nacional de proteção de dados (Autorização 5985/2014). Os indivíduos convidados a participar neste estudo assinaram uma declaração e um consentimento informado. Foi também realizado um formulário clínico, com informações sobre os indivíduos, relevantes para o estudo, preenchido pelas equipas médicas dos Centros Pneumológicos da Venda Nova e de Almada-Seixal

Para este trabalho foram selecionados cinquenta e quatro indivíduos saudáveis, IGRA negativos, vinte indivíduos com tuberculose pulmonar e doze indivíduos com tuberculose extrapulmonar. Os indivíduos pertencentes ao grupo dos doentes foram atendidos nos Centros de Diagnóstico Pneumológico (CDP) da Venda Nova e de Almada-Seixal. As colheitas de sangue dos indivíduos saudáveis, pertencentes ao grupo controlo foram realizadas no Instituto de Tecnologia Química e Biológica António Xavier (ITQB NOVA) sob a orientação do CDP da Venda Nova.

Os indivíduos foram selecionados para este estudo com base em alguns critérios previamente definidos:

- Com idades compreendidas entre os 18 e os 65 anos;
- Indivíduos sem nenhuma outra infeção respiratória (no caso dos indivíduos com tuberculose);
- Indivíduos HIV negativos;
- Indivíduos que não tenham sido transplantados;
- Indivíduos que não tenham diabetes;
- Indivíduos que não tenham falha renal crónica.

As amostras com sinais de hemólise ou com um resultado do teste do IGRA indeterminado foram excluídas do estudo.

3.3 Recolha das amostras de soro sanguíneo

Para o grupo dos doentes foram recolhidos cerca de 7 ml de sangue, enquanto que para o grupo dos controlos foram recolhidos 10,5 ml de sangue (7 ml para o estudo de metabolómica e cerca de 3,5 ml para o ensaio IGRA)

A colheita de sangue para o ensaio de metabolómica foi efetuada em tubos de ativação da coagulação (S-Monovette, Z-gel, Sarstedt, Ref: 01.1602.001). Estes tubos são revestidos com partículas de sílica micronizadas que causam a rápida coagulação do sangue, contendo também uma barreira de gel capaz de separar eficazmente o soro das células sanguíneas e do fibrinogénio, prevenindo também as trocas entre as células sanguíneas e o soro.

Para o ensaio de IGRA (QuantiFERON® – TB Gold IT, QIAGEN), foram colocados cerca de 1ml de sangue em cada tubo, sendo estes tubos específicos para o ensaio do IGRA.

As amostras para ensaio IGRA e para ensaios de metabolómica foram transportadas à temperatura ambiente e à temperatura de 4°C, respetivamente, para o Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA), juntamente com as declarações, os consentimentos informados e os formulários clínicos dos indivíduos.

No laboratório de Micobactérias do INSA procedeu-se à separação do soro (complementar com a informação no protocolo), à descontaminação do soro e à realização do ensaio do IGRA para o grupo dos controlos.

Para o estudo de metabolómica, o sangue foi deixado a coagular por três horas a 4°C após a sua recolha. De seguida o sangue foi centrifugado a 3000 rpm a 4°C, durante trinta minutos. O soro recolhido após a centrifugação passou por filtros de 0,2 µm (Sterile Acrodisc, syringe filters with Supor membrane, 32 mm) para remover as micobactérias presentes no soro. Por fim, o soro filtrado foi dividido em alíquotas de 500 µl e armazenado a -80°C. Para os testes do IGRA foram utilizados kits Quilaban. As amostras foram transportadas do INSA para o ITQB num contentor de azoto líquido e armazenadas a -80°C.

3.4 Preparação das amostras para NMR

Antes da aquisição dos espectros de NMR, foi necessário remover macromoléculas, tais como proteínas e, também os componentes lipídicos do soro sanguíneo por ultrafiltração.

Antes da ultrafiltração das amostras foi necessário realizar lavagens sucessivas dos filtros. Este passo serviu para remover a camada de glicerol da membrana dos filtros. Assim foi necessário lavar os filtros três vezes com 500 μ l de uma solução de NaCl a 150 mM e outras três vezes com 500 μ l de água milliQ. Para cada lavagem procedeu-se à centrifugação dos filtros a uma rotação de 15000 x g a 4°C, durante cerca de 10 minutos. Posteriormente à lavagem dos filtros, efetuou-se a ultrafiltração das amostras, com a centrifugação a 15000 x g a 4°C durante quarenta e cinco minutos. Uma hora antes da filtração do soro, este foi descongelado a 4°C em gelo.

Foram colocados 400 μ l de soro filtrado em tubos de NMR (ALDRICH®528-PP, 5mm). Para as amostras cujo o volume, resultante da filtração, foi inferior a 400 μ l, foi necessário adicionar-se água para obter-se um volume de 400 μ l. Aos tubos de NMR adicionou-se cerca de 160 μ l de solução contendo 50 mM de uma solução tampão fosfato a pH 7.0 e 0.25 mM DSS em água deuterada de modo que todas as amostras apresentassem um volume final de 560 μ l. A água deuterada age como solvente para estabilizar a aquisição de dados por NMR. O espectrómetro de NMR “trava” na frequência de ressonância do deutério, a qual não é observada, permitindo que seja atingida a estabilidade da frequência do campo. Isto vai evitar o aparecimento de sinais amplos e descaracterizados. Por outro lado, o tamponamento do pH vai permitir que não haja variações nos desvios químicos dos picos espectrais.

Após se obter o volume de 560 µl nos tubos de NMR procedeu-se à homogeneização por inversão total, tendo-se cuidado para evitar a formação de bolhas.

3.5 Aquisição de espectros de NMR

Os espectros foram recolhidos no Centro de Ressonância Magnética António Xavier – CERMAX, situado no ITQB a partir de um espectrómetro de 800 MHz (Bruker BioSpin, Rheinstetten, Germany), obtendo-se espectros 1D (NOESY e CPMG) e 2D (J-resolved e TOCSY) para se proceder à identificação e quantificação de metabolitos. Utilizou-se a sonda TXI-Z H-C/N/-D (5mm) com deteção inversa (otimizada para a observação de ^1H) projetada para operar a três frequências fixas (^1H , ^{13}C e ^{15}N), com gradientes de campo pulsado no eixo Z e sintonização automática de frequências (Anon 2007).

Os espectros de NMR de protão de uma dimensão foram obtidos num espectrómetro de 800 MHz, a uma temperatura de 298k, operando a uma frequência de 800,33 MHz. Os espectros foram adquiridos por uma frequência de pulso “*noesygppr1d*” (Bruker BioSpin, Rheinstetten, Germany). Os parâmetros de aquisição dos espectros NOESY estão descritos na Tabela 3.

Tabela 3: Parâmetros de aquisição de espectros 1D NOESY

Experiência	1D NOESY
Sequência de pulso	<i>noesygppr1d</i>
Tamanho do campo (TD)	65536 = 64k
Número de scans (NS)	128
Largura do espectro (SW)	30.04 ppm
Tempo de aquisição (AQ)	1,36 s
Tempo de mistura (d8)	0,01 s
Tempo de relaxação (d1)	4 s

Além da aquisição de espectros 1D, foram também adquiridos espectros 2D. Os parâmetros de aquisição dos espectros com as sequências de pulso J-resolved (Bruker BioSpin, Rheinstetten, Germany) e ^1H - ^1H TOCSY (Bruker BioSpin, Rheinstetten, Germany) estão descritos na Tabela 4

Tabela 4: Parâmetros de aquisição de espectros bidimensionais

Experiência	Parâmetros de aquisição	
	J-resolved	^1H - ^1H TOCSY
Sequência de pulso	“jresgpprqf”	“dipsi2esgpph”
Número de scans, NS	32	16
Tempo de relaxação entre scans (s)	2	1,5
Tempo de mistura (ms)	30	30
Número de pontos de tamanho do FID na 1ª dimensão	8192	400
Número de pontos de tamanho do FID na 2ª dimensão	40	2048
Largura do espectro na 1ª dimensão (ppm)	0.0976	12
Largura do espectro na 2ª dimensão (ppm)	16.6154	16.019

3.6 Processamento dos espectros de NMR

Para se conseguir obter uma boa identificação e quantificação de metabolitos, é necessário a realização do processamento dos espectros pelo TopSpin 3.2 (Bruker BioSpin Ltd., Canada). Este passo envolve várias etapas:

- . Transformação de Fourier
- . Correção da fase automática ou manual quando necessário
- . Correção da linha de base automática ou manual quando necessário
- . Ampliação da linha do espectro (LB: 1 Hz)

3.7 Identificação e quantificação de metabolitos

A identificação dos metabolitos presentes nas amostras de soro sanguíneo foi realizada como auxílio do software Chenomx NMR Suite 8.1 (Chenomx Inc., Edmonton, AB, Canada) (Weljie et al. 2006). Este software apresenta uma base de dados interna, que contém os espectros individuais de metabolitos adquiridos com a sequência de pulso *noesypr*, num espectrómetro de 800MHz. A identificação dos metabolitos foi também efetuada através das bases de dados do metaboloma humano (HMDB 3.6) (Wishart et al. 2013) e do banco biológico de ressonância magnética (BMRB) (Ulrich et al. 2008). Os espectros 2D adquiridos, foram também uma mais valia na identificação de metabolitos. Estes foram usados para identificar metabolitos que apenas apresentassem um pico no espectro 1D ou que apresentassem sinais similares com outros metabolitos.

O cálculo da concentração metabólica foi realizado com base na concentração conhecida do composto de referência, neste caso o DSS. A concentração de DSS na amostra foi calculada previamente de acordo com a diluição da amostra, através da equação A, representada em baixo.

Equação A: Equação usada no cálculo da concentração de DSS após diluição da amostra

$$[\text{DSS}]_{\text{aparente}} = \frac{[\text{DSS}]_{\text{inicial}} \times \text{Volume do buffer}}{\text{Volume do soro}}$$

3.8 Binning

Para a análise *fingerprinting* foi utilizado o *binning* de maneira a reduzir a dimensão dos dados. Esta técnica permite dividir o espectro em regiões de igual tamanho que iram ser integradas para obter as áreas espectrais dos *bins*. Esta técnica foi conseguida através da utilização do software Chenomx NMR Suite 8.1 (Chenomx Inc., Edmonton, AB, Canada) (Weljie et al. 2006). Os dados do *binning* foram normalizados com a área total

de cada espectro, através da exclusão da ressonância da água. Além do sinal da água foram excluídos os sinais da ureia, devido às trocas iônicas deste metabolito com a água e removidos os sinais correspondentes a contaminantes, como a acetona e o etanol, assim como os sinais de DSS. Foram também removidos os sinais referentes ao glicerol, dado que este se encontra presente na membrana dos filtros utilizados para a ultrafiltração. Adicionalmente foram excluídas as regiões do espectro referentes aos princípios ativos utilizados no tratamento da tuberculose, assim como as regiões de ruído espectral.

Foi realizado o *binning* para a linha de espectro, utilizando vários tamanhos de *bins*. O *binning* realizado para a linha de subtração foi utilizado para o cálculo da área espectral por identificar e já identificada, sendo que a linha de subtração corresponde à linha espectral obtida após identificação e quantificação dos metabolitos.

3.9 Processamento dos dados para análise estatística

Antes da realização da análise estatística, os dados obtidos foram organizados em duas matrizes distintas. Nestas matrizes as linhas correspondem às amostras e as colunas compreendem as variáveis. No caso da matriz usada para a análise *fingerprinting*, as variáveis correspondem à soma das intensidades dos sinais no intervalo de *bin* e no caso da análise do perfil metabólico, as variáveis correspondem às concentrações metabólicas dos metabolitos identificados e quantificados.

Os dados utilizados para a análise multivariada e da análise univariada foram importados para o SIMCA-P 13.0.3 e para o MetaboAnalyst 3.0, respetivamente, onde foi utilizada a unidade de variância como método de escalamento.

No caso da análise do perfil não foram utilizados na análise estatística, os metabolitos correspondentes à ureia, ao etanol, à acetona, ao glicerol e ao DSS. Além

disso, os metabolitos que foram identificados, mas não quantificados também não foram utilizados na análise estatística.

3.10 Análise Univariada

A análise univariada dos dados obtidos através da análise do perfil metabólico foi realizada no MetaboAnalyst 3.0 (www.metaboanalyst.ca).

Para esta análise, a significância estatística de cada variável (concentração de metabolitos) foi determinada em indivíduos com tuberculose e em indivíduos saudáveis utilizando o teste *t-Student*, com o teste de Fisher. Para este teste foi considerado significativo um valor de p inferior a 0.05, sendo traçados os gráficos de bigodes para os metabolitos com valor de p inferior a 0.05. Foi ainda calculada a taxa de descoberta de falsos positivos (FDR) através do MetaboAnalyst.

3.11 Análise Multivariada

A análise multivariada foi realizada com recurso ao software SIMCA 13.0.3 (Umetrics, Umeå, Sweden).

O conjunto de dados foi analisado, através do método não supervisionado, PCA e do método supervisionado, PLS-DA. No PCA não é tida em conta qualquer informação prévia sobre os dados, enquanto no PLS-DA é dada a informação à cerca das classes onde as amostras estão inseridas. Os gráficos de *scores* de ambos os modelos representam um resumo da variação de todas as amostras, mostrando a relação entre elas. O PCA permite-nos também detetar *outliers*. No PLS-DA, o gráfico de *loadings* permite identificar as variáveis responsáveis pela separação das amostras.

3.12 Validação dos modelos de análise multivariada

Para se ter confiança estatística para validar o modelo de PCA são requeridos valores de R^2 e de Q^2 superiores a 0,5. A diferença entre os valores destes dois parâmetros não deve ultrapassar os 0,3. O parâmetro *hotelling* T^2 permite verificar a presença de *outliers* no conjunto de dados.

Para o modelo de PLS-DA, são utilizados os parâmetros R^2X , R^2Y e Q^2 . Para estes parâmetros são aconselháveis valores maiores que 0,5. Neste modelo foi também utilizado o valor de p do teste ANOVA da validação-cruzada para determinar a significância estatística da separação entre os grupos (Triba et al. 2015; Umetrics 2011). Foram realizados testes de permutação para validar a qualidade do modelo, consistindo na permutação aleatória de membros das classes, repetindo o processo por 200 vezes.

Para escolher as variáveis de interesse são usados os valores de VIP superiores a 1.

3.13 Análise das vias metabólicas

De modo a serem identificadas as vias metabólicas que mais são afetadas pela infecção por *Mtb*, utilizou-se a análise de enriquecimento de conjuntos de metabolitos (MSEA) (Xia & Wishart 2010). Esta análise foi realizada utilizando o site MetaboAnalyst 3.0 (www.metaboanalyst.ca) com recurso aos protocolos usados para este servidor (Xia & Wishart 2011; Xia et al. 2016).

Capítulo IV: Resultados e Discussão

4.1 Caracterização dos voluntários

Foi realizada uma análise do perfil dos indivíduos que participaram neste estudo, com base em características como o género, a idade, o índice de massa corporal e os seus hábitos tabágicos.

Como se pode observar pelo gráfico de barras representado pela figura 10, no grupo controlo, existe uma distribuição de indivíduos, com base no género, mais ou menos equivalente. Já em relação ao grupo de doentes pulmonares e de doentes extrapulmonares, a distribuição dos indivíduos por género não é homogénea. Foi possível obter mais amostras de indivíduos do género masculino do que de indivíduos do género feminino, no grupo de doentes pulmonares e vice-versa no grupo de doentes extrapulmonares. A distribuição de doentes por género está de acordo com estudos epidemiológicos publicados (García-rodríguez et al. 2011; Sunnetcioglu et al. 2015; Paper 2015).

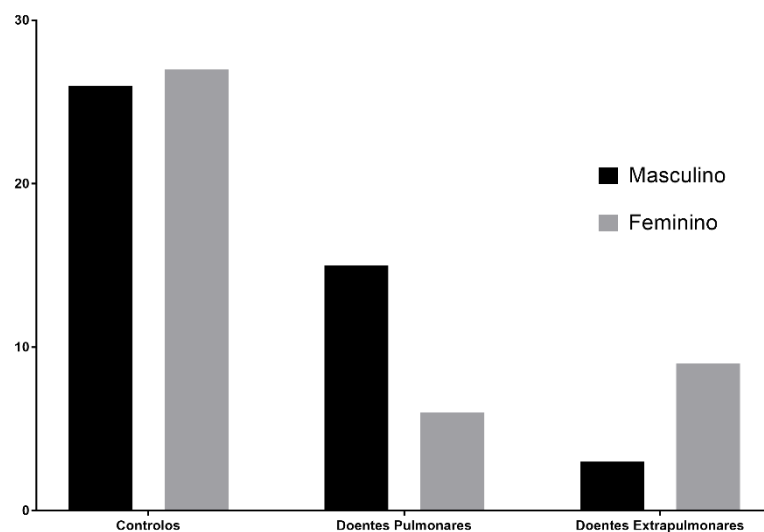


Figura 10: Gráfico de barras representativo da distribuição dos indivíduos saudáveis, com tuberculose pulmonar e com tuberculose extrapulmonar por género.

Em relação à idade dos indivíduos podemos verificar que a grande maioria dos doentes com tuberculose pulmonar tem idades inferiores a 50 anos. No caso da tuberculose extrapulmonar, a maioria dos indivíduos apresentava idades superiores a 50 anos (Figura 11).

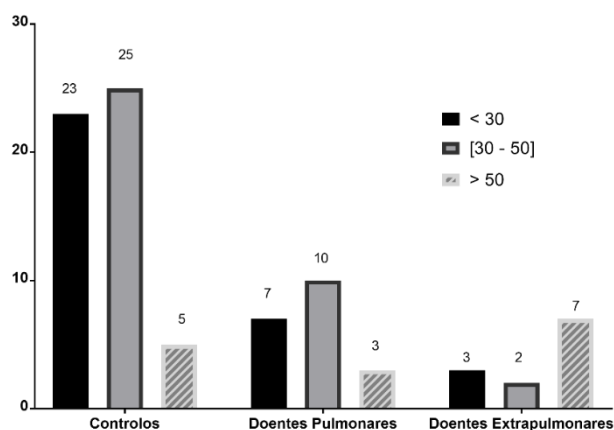


Figura 11: Gráfico de barras representativo da distribuição dos indivíduos saudáveis, com tuberculose pulmonar e com tuberculose extrapulmonar pela idade. < 30 anos; 30 a 50 anos; > 50 anos

Relativamente ao índice de massa corporal verifica-se que a maioria dos indivíduos saudáveis apresentam um IMC normal (Figura 12). Aproximadamente 21% da população saudável estudada apresenta excesso de peso. Nota-se que os indivíduos com tuberculose pulmonar estão distribuídos mais ao menos proporcionalmente entre os índices de massa corporal baixo e normal. Já nos indivíduos com tuberculose extrapulmonar, os indivíduos distribuem-se entre os índices de massa corporal normal e elevado, quase na mesma proporção. Esta distribuição de doentes, segundo o seu IMC, está de acordo com estudos epidemiológicos já publicados (Leung et al. 2015)

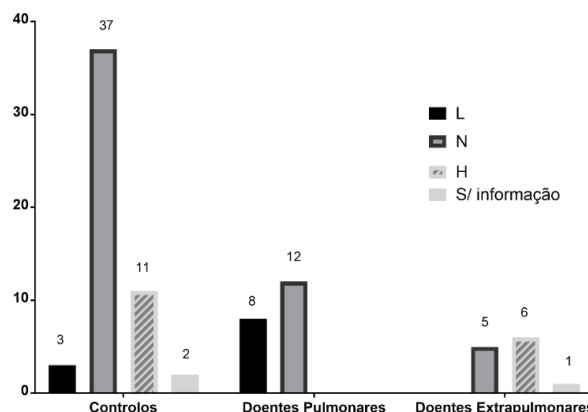


Figura 12: Gráfico de barras representativo da distribuição dos indivíduos saudáveis, com tuberculose pulmonar e com tuberculose extrapulmonar segundo o seu IMC. L: IMC baixo; N: IMC normal; H: IMC elevado

Como podemos observar pelo gráfico de barras representado pela figura 13, a maioria dos indivíduos saudáveis não fuma. Já no grupo de indivíduos com tuberculose pulmonar os indivíduos estão distribuídos mais ou menos na mesma proporção pelas três categorias (fumadores, não fumadores e ex-fumadores). Por outro lado, no grupo de indivíduos com tuberculose extrapulmonar, a maioria dos indivíduos não fuma.

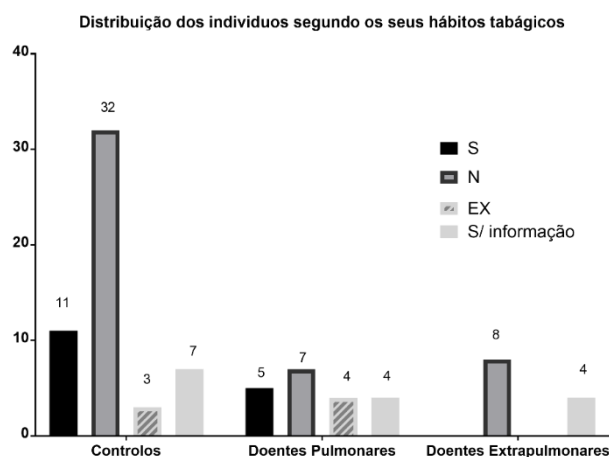


Figura 13: Gráfico de barras representativo da distribuição dos indivíduos saudáveis, com tuberculose pulmonar e com tuberculose extrapulmonar segundo os seus hábitos tabágicos. S: fumador; N: não fumador; EX: ex-fumador

4.2 Otimização do protocolo para aquisição de espectros de NMR

O soro sanguíneo é um fluido biológico com grande complexidade, nomeadamente devido ao elevado número de constituintes e à sua diversidade química. Esta complexidade resulta em espectros de ^1H NMR unidimensionais, como o espectro A representado na figura 14. Este espectro apresenta muitos sinais amplos, característicos da presença de macromoléculas, sobrepostos com sinais estreitos, correspondentes aos sinais dos metabolitos. De modo a conseguir estender o número de metabolitos detetados e de possibilitar a sua quantificação, a partir destes espectros, foi necessário remover os sinais das proteínas e de outras macromoléculas. Antes do início deste trabalho foram testadas várias técnicas para remover estes sinais e foram comparados os perfis de espectros adquiridos com duas sequências de pulso comumente utilizadas (CPMG e 1D NOESY). Adicionalmente compararam-se os perfis destes espectros com os obtidos para amostras de soro não tratadas e para as quais foi realizada, por ultrafiltração, uma prévia remoção de macromoléculas com massa superior a 5kDa.

No espectro B, adquirido com uma sequência de pulso CPMG e sem recurso à ultrafiltração, é possível perceber que os sinais das proteínas, de relaxação rápida foram removidos quase na totalidade. Neste espectro nota-se a presença de alguns sinais largos correspondentes a macromoléculas de relaxação lenta, como alguns lípidos. Tanto nos espectros CPMG (espectro C), como no espectro NOESY (espectro D) adquiridos após a realização de ultrafiltração, é possível observar apenas os sinais correspondentes aos metabolitos. No entanto os sinais dos metabolitos apresentam uma maior intensidade para o espectro NOESY.

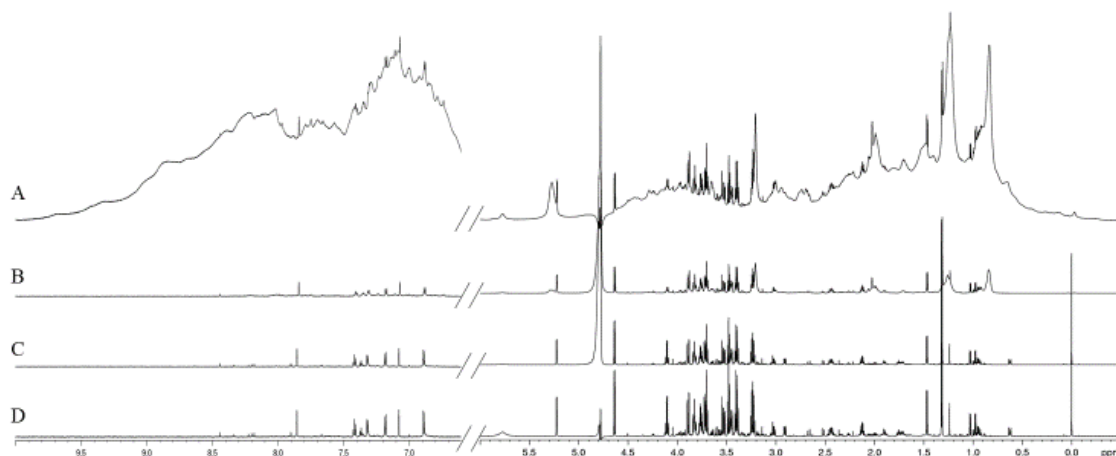


Figura 14: Representação dos espectros NOESY e CPMG de uma amostra controle. **A:** espectro 1D NOESY sem prévia ultrafiltração da amostra; **B:** Espectro CPMG sem prévia ultrafiltração da amostra; **C:** Espectro CPMG com prévia ultrafiltração da amostra; **D:** Espectro 1D NOESY com prévia ultrafiltração da amostra

Optou-se então pela utilização do processo de ultrafiltração e aquisição de espectros NOESY, dado neste caso ser obtida uma maior taxa de recuperação de metabolitos. Esta informação é corroborada por estudos já publicados (Daykin et al. 2002; Tiziani et al. 2008). Além disso, optou-se pela utilização de espectros adquiridos com a sequência de pulso NOESY dado que estes produzem uma melhor supressão do sinal do solvente, neste caso a água. Além disso, os espectros NOESY melhoram a quantificação da concentração dos metabolitos.

Na mesma ocasião foram realizados ensaios para determinar a estabilidade da amostra. Para isso adquiriram-se espectros com sequência de pulso NOESY a diferentes períodos de tempo após ultrafiltração para uma mesma amostra. Foi adquirido um espectro logo após a preparação da amostra, e outros cinco espectros após quatro, nove, treze, dezasseis e vinte e sete horas após a preparação da amostra. Comparando os espectros no software TopSpin 3.2 verificou-se que ao longo do tempo ocorrem alterações no pH da amostra. Esta variação do pH mostrou-se visível nos sinais da histidina, metabolito extremamente sensível a variações no pH. Além disso observou-se que alguns sinais começavam a deteriorar-se após 3 horas da preparação da amostra.

Tendo sido detetada a presença dos princípios ativos, usados no tratamento da tuberculose, em soros de doentes, foi necessário adicionar à estratégia de análise de dados por *fingerprinting* a análise do perfil metabólico.

Uma vez que apenas estava estipulado o uso da análise *fingerprinting*, optou-se por utilizar uma sequência de pulso NOESY da Bruker, em vez da sequência de pulso NOESY estipulada pelo Chenomx. Realizando-se a identificação e quantificação metabólica no Chenomx verificou-se que alguns picos de metabolitos não encaixavam com os picos da base de dados do Chenomx, tornando este processo ligeiramente complicado. Este problema foi resolvido com recurso a espectros 2D.

4.3 Desconvolução dos espectros de ^1H NMR de amostras de soro de indivíduos saudáveis e doentes

Os espectros de ^1H NMR das amostras de soro sanguíneo, apresentam várias dezenas de sinais que resultam de dezenas de metabolitos. Como se pode observar através da figura 15 e da tabela 5, foi realizada a identificação de sessenta e um metabolitos através dos espectros de NMR uni e bidimensionais. Esta identificação, foi realizada com o auxílio de várias bases de dados, nomeadamente a base de dados do metaboloma humano (HMDB), o banco de dados biológicos de ressonância magnética (BMRB) (<http://www.bmrbl.wisc.edu/>) e a base de dados interna do Chenomx NMR Suite. Foram ainda usados nesta identificação artigos publicados (Psychogios, David D. Hau, et al. 2011; Tiziani et al. 2008).

Espectros de NMR bidimensionais, nomeadamente ^1H - ^1H TOCSY e J-resolved foram adquiridos para as amostras de ambos os grupos (controlos e doentes), ajudando na identificação de vários metabolitos. A título de exemplo, foram identificados através do espectro ^1H - ^1H TOCSY metabolitos como o 2-aminobutirato, o 3-hidroxisobutirato, a fructose e o propilenoglicol, como ilustrado nas figuras B1 a B4 do anexo II. As figuras C 1 e C 2 do anexo III, representativas de regiões de espectros J-resolved também foram usadas para a identificação de alguns sinais de metabolitos com mais de um pico.

Os sessenta e um metabolitos identificados foram enumerados conforme mostrado na tabela 5 e detalhado nas figuras A1-A7 do anexo I. As principais classes de metabolitos detetadas correspondem a aminoácidos, ácidos orgânicos, glícidos e intermediários nucleosídeos (Tabela 19 do anexo IV)

Para determinar a área espectral identificada definiu-se um *bin* único com a linha do espectro e com a linha de subtração do espectro. A linha de subtração do espectro obteve-se após identificação e quantificação dos metabolitos, correspondendo aos sinais espectrais que não foram identificados. Assim, verificou-se que a área espectral identificada correspondeu a cerca de 93,4% do espectro, deixando apenas 6,6% do espectro por identificar.

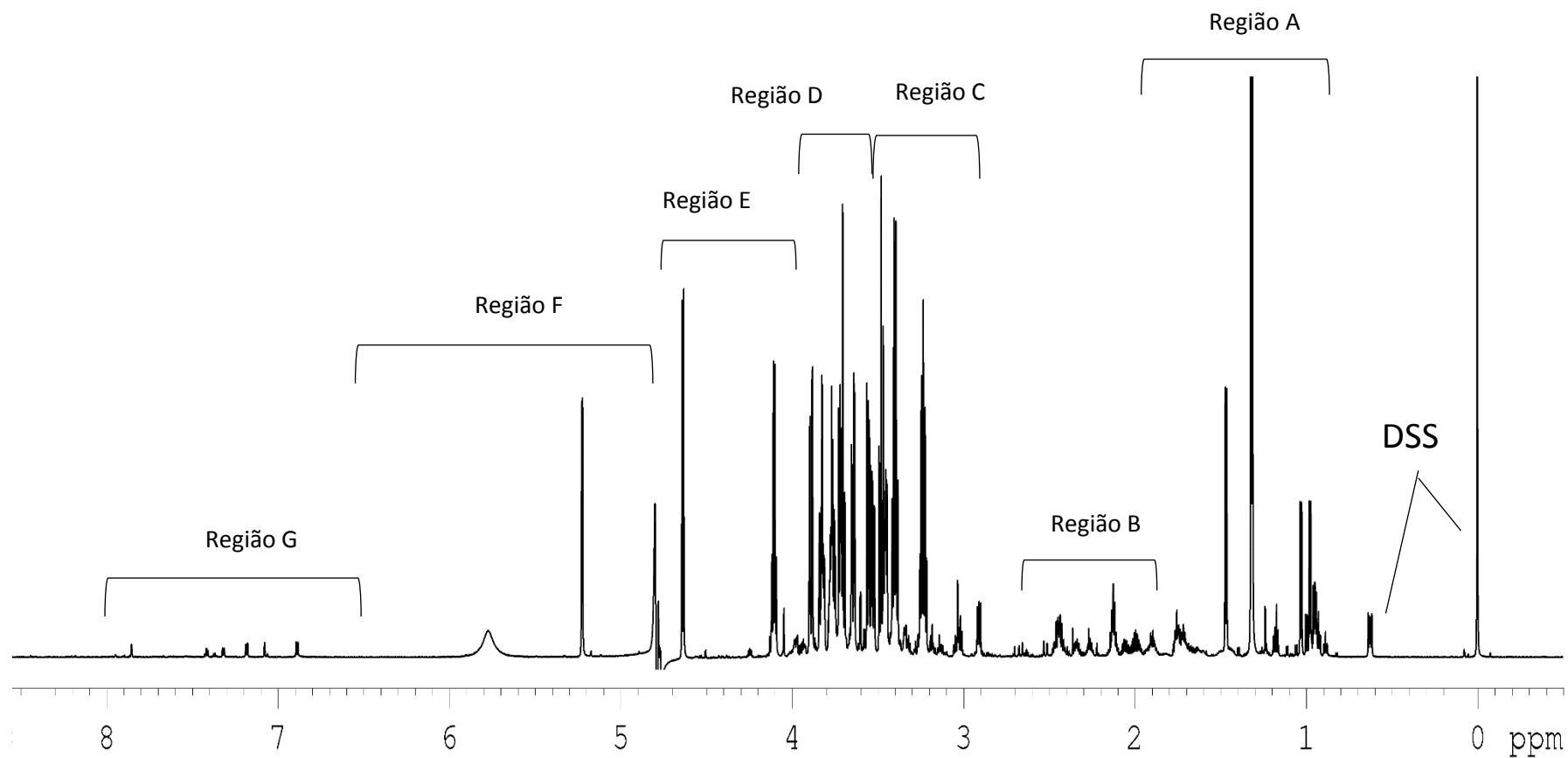


Figura 15: Espectro 1D NOESY de uma amostra do grupo controle, adquirido com prévia ultrafiltração da amostra. As regiões A a G apresentadas, estão detalhadas nas figuras A1-A7 do anexo I.

Os valores das concentrações metabólicas obtidos para o grupo controle e de doentes foram calculados com base no valor da concentração de DSS nas amostras, com base na equação A apresentada no capítulo Materiais e Métodos. Os valores médios e do desvio padrão de cada metabolito, dentro de cada grupo experimental, são apresentados na tabela 5.

De forma a avaliar a quantificação realizada, foram comparados os valores médios e os desvios padrão das concentrações metabólicas dos indivíduos controle com as concentrações descritas nas bases de dados.

Nem todas as concentrações metabólicas encontradas no soro, calculadas a partir do NMR, estão de acordo com os valores da base de dados. Cinquenta e dois metabolitos apresentam concentrações já descritas nas bases de dados. Destes cinquenta e dois metabolitos, apenas trinta e três exibiram uma boa concordância com os valores descritos na base de dados. Ou seja, trinta e três metabolitos apresentaram valores médios de concentração dentro do desvio padrão dos valores da base de dados. Além disso, cerca de seis metabolitos têm concentrações médias mais elevadas, relativamente às descritas (lactato, arginina, frutose, malonato, uridina e inosina) enquanto treze metabolitos têm concentrações médias menores, relativamente às descritas (lisina, serina, citrato, histidina, fenilalanina, ascorbato, betaína, cisteína 3-hidroxisobutirato, formato, hipoxantina, propileno glicol e triptofano)

Pela figura 16 é possível verificar que o metabolito 1, correspondente à glucose, apresenta valores de desvio padrão ligeiramente acima dos valores descritos na base de dados. Como se pode observar pelas figuras 17 e 18, existem alguns metabolitos, nomeadamente a frutose (16), o acetato (24), o malonato (39), o 2-hidroxisovalerato (46) e a inosina (53) que apresentam uma maior discrepância em termos dos valores de desvio padrão das suas concentrações. No caso dos metabolitos 16, 24 e 46 verificou-se que a esta enorme discrepância se deve à presença de *outliers* no conjunto de amostras do grupo controle.

O efeito do tamanho da população e as diferentes técnicas de extração de metabolitos usadas, podem também ser duas das causas associadas aos diferentes valores quantificados (Psychogios, David D. Hau, et al. 2011).

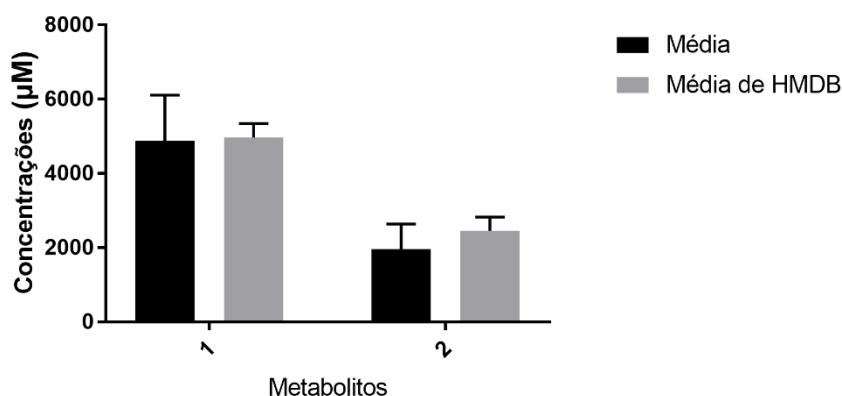


Figura 16: Histograma ilustrativo da comparação das concentrações metabólicas quantificadas por NMR, em amostras de indivíduos saudáveis, e descritas na literatura, relativamente aos metabolitos 1 e 2.

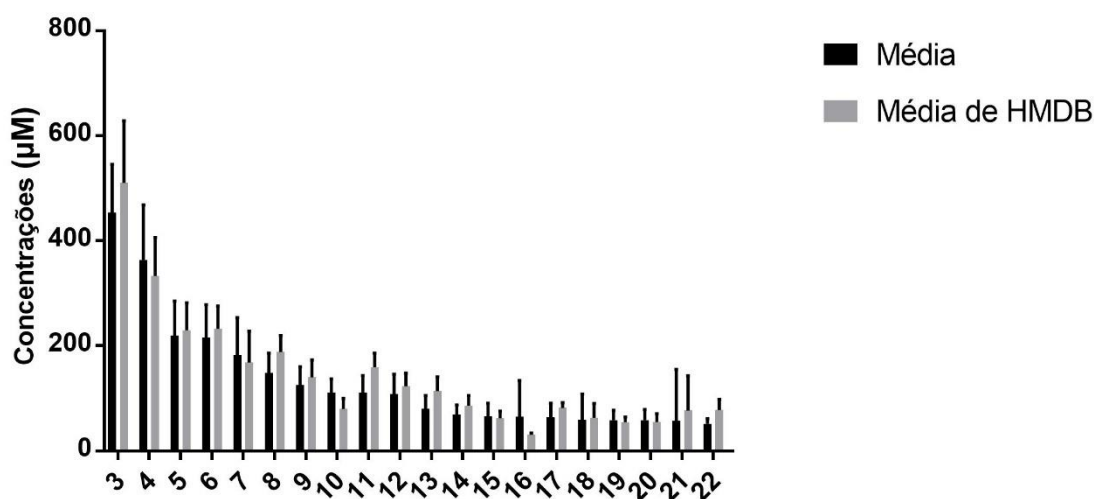


Figura 17: Histograma ilustrativo da comparação das concentrações metabólicas quantificadas por NMR, em amostras de indivíduos saudáveis, e descritas na literatura, relativamente aos metabolitos 3 a 22

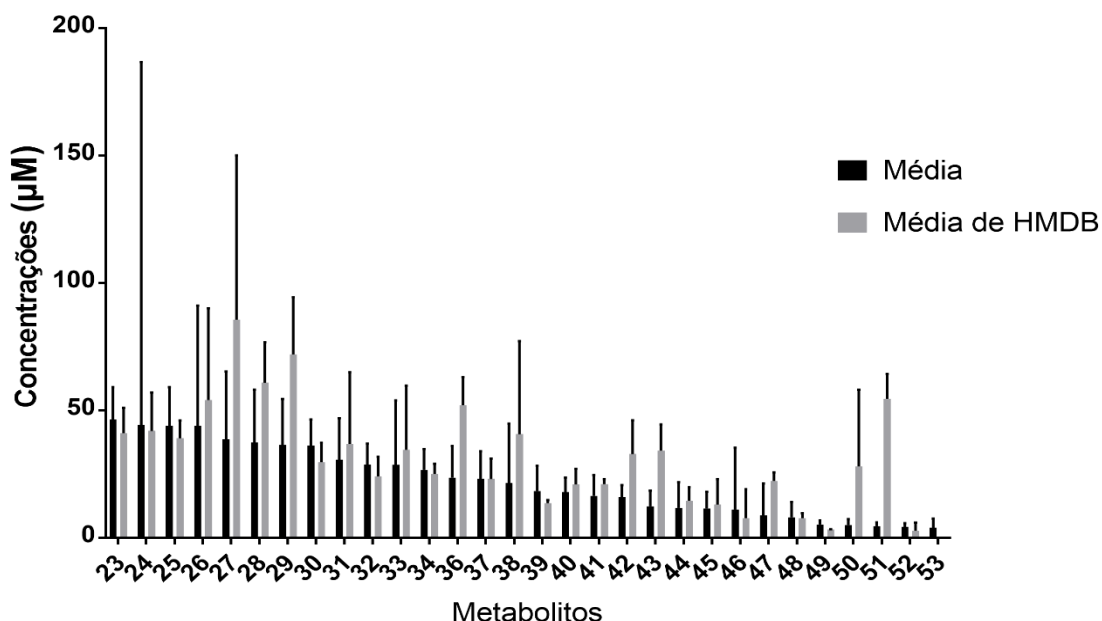


Figura 18: Histograma ilustrativo da comparação das concentrações metabólicas quantificadas por NMR, em amostras de indivíduos saudáveis, e descritas na literatura, relativamente aos metabólitos 23 a 34 e 36 a 53.

Com base nos valores indicativos da tabela 5 é possível constatar que existem algumas diferenças na quantificação realizada neste estudo e nas quantificações descritas na base de dados. No caso da fucose, não foram encontrados valores descritos, relativamente à sua quantificação no sangue. Para o etanol, a acetona, o glicerol, o 1,5-anidro sorbitol, a cafeína e a ureia, não foi realizada a sua quantificação. Isto porque, o glicerol se encontra presente nas membranas do filtro usado na ultrafiltração das amostras. Para o 1,5-anidro sorbitol, apenas foi possível fazer a sua identificação. Isto porque, a base de dados do Chenomx não permitiu calcular a sua concentração absoluta. Para algumas amostras foi detetada a presença de sinais de etanol e de acetona bastante intensos. Este facto resultou de contaminações ocorridas durante a manipulação das amostras. A concentração da cafeína no soro vai depender do número de horas que o indivíduo está sem comer antes de se efetuar a recolha de sangue e dos hábitos alimentares do mesmo.

Através da tabela 5 é também possível verificar algumas diferenças, em termos das concentrações metabólicas, entre as amostras do grupo controlo e as amostras do grupo de doentes.

	Metabolito	Controlos (n:54)	Doentes (n:32)	Literatura
1	Glucose	4880,9 (3651,5 – 6110,3)	6379.1 (2000,3 – 10757,9)	4971.3 (4598,5 – 5344,1)
2	Lactato	1955 (1276,6 – 2633,4)	2453.5 (1491,2 – 3415,8)	1489.4 (1118,2 – 1860,6)
3	Glutamina	453.6 (361,3 – 545,9)	503.9 (330,5 – 677,7)	510,4 (392,2 – 628,6)
4	Alanina	363.8 (259,4 – 468,2)	391.6 (236,3 – 546,9)	333 (259 – 407)
5	Glicina	219.4 (153 – 285,8)	245.7 (170,4 – 321)	230 (178 – 282)
6	Valina	215.8 (153,1 – 278,5)	220.5 (153,7 – 287,3)	233 (190 – 276)
7	Prolina	182.6 (111,4 – 253,8)	185.8 (111,9 – 259,7)	168.0 (108.0-228.0)
8	Lisina	148.8 (111,6 – 186)	138.8 (94 – 183,6)	188.0 (156.0-220.0)
9	Treonina	125.8 (91,5 – 160,1)	119.3 (68,8 – 169,8)	140 (107 – 173)
10	Arginina	111.5 (85,8 – 137,2)	114 (91,1 – 136,9)	80 (60.0 - 100)
11	Serina	111.4 (79,3 – 143,5)	132.3 (98,9 – 165,7)	159.8 (133,2 – 186,4)
12	Leucina	108.3 (70,5 – 146,1)	115.8 (77,5 – 154,1)	123 (98 – 148)
13	Citrato	79.7 (53,9 – 105,5)	68.5 (48 – 89)	114.2 (87,2 – 141,2)
14	Creatinina	69.3 (51,2 – 87,4)	75.4 (43,3 – 107,5)	86.6 (67,8 – 105,4)
15	Isoleucina	66.4 (42 – 90,8)	70.6 (45,8 – 95,4)	62 (48 – 76)
16	Frutose	64.8 (14,4 – 300,1)	89.6 (0 – 183,5)	31 (28-34)
17	Histidina	63.9 (36,7 – 91,1)	62.1 (37,9 – 86,3)	82 (72 – 92)
18	Cistina	58.9 (9,5 – 108,3)	46.8 (29,2 – 64,4)	62.9 (35,1 – 90,7)
19	Tirosina	58.5 (39,4 – 77,6)	65.3 (43,4 – 87,2)	54.5 (44,8 – 64,2)
20	Ornitina	58.3 (37,6 – 79)	78.7 (45,7 – 111,7)	55 (39 – 71)
21	3-Hidroxibutirato	57.5 (46,4 – 68,4)	47.8 (0 – 150,9)	76.9 (10,6 – 143,2)
22	Fenilalanina	51.2 (41,1 – 61,3)	53.5 (39,9 – 67,1)	78.1 (57,6 – 98,6)
23	Asparagina	46.4 (33,7 – 59,1)	39.9 (24,6 – 54,6)	41.0 (31.0-51.0)
24	Acetato	44.3 (20,1 – 159,6)	26.5 (11,1 – 41,9)	41.9 (26,8 – 57)
25	Manose	43.9 (28,7 – 59,1)	65.4 (30,6 – 100,2)	39 (32 – 46)
26	2-Hidroxibutirato	43.9 (22,6 – 364,6)	35.7 (19 – 52,4)	54 (8 – 80)
27	Glutamato	38.6 (12 – 65,2)	53 (34,4 – 71,6)	85.5 (21 – 150)
28	Ascorbato	37.4 (16,7 – 58,1)	34.8 (18,6 – 51)	60.8 (44,9 – 76,7)
29	Betaina	36.4 (18,5 – 54,3)	40.5 (25,2 – 55,8)	72 (49,6 – 94,4)
30	Carnitina	36.2 (26 – 46,4)	38.5 (26,7 – 50,3)	29.74 (22,19 – 37,3)
31	Creatina	30.7 (14,4 – 47)	30 (11,9 – 48,1)	36.7 (8,4 – 65)
32	mio-Inositol	28.7 (20,5 – 36,9)	32.6 (21,5 – 43,7)	24 (16,2 – 31,8)
33	Piruvato	28.7 (3,5 – 53,9)	40.8 (3,1 – 78,5)	34.5 (9,3 – 59,7)
34	Metionina	26.6 (18,4 – 34,8)	22.1 (14 – 30,2)	25.0 (21.0-29.0)
35	Fucose	25.1 (18,9 – 31,3)	23.6 (15,2 – 32)	NE
36	Cisteína	23.5 (11 – 36)	18.4 (10,3 – 26,5)	52 (41 – 63)
37	2-Aminobutirato	23.0 (12,1 – 33,9)	23.6 (13,6 – 33,6)	23.0 (15.0-31.0)
38	Acetoacetato	21.4 (18,4 – 168,8)	26.4 (1,1 – 51,7)	40.6 (4,1 – 77,1)
39	Malonato	18.3 (8,2 – 28,4)	15.6 (5,4 – 25,8)	13.5 (12,3 – 14,7)
40	Aspartato	17.9 (12,2 – 23,6)	21.5 (13,6 – 29,4)	20,9 (14,8 – 27)
41	3-Hidroxisobutirato	16.4 (8,1 – 24,7)	12.6 (6 – 19,2)	21 (19 – 23)
42	Formato	15.9 (11,2 – 20,6)	33.3 (0 – 120,9)	32,8 (19,5 – 46,1)
43	Hipoxantina	12.2(6 – 18,4)	20.3 (9 – 31,3)	34,2 (23,9 – 44,5)

44	Colina	11.7 (1,6 – 21,8)	10.4 (6,4 – 14,4)	14,5 (9,2 – 19,8)
45	trans-4-Hidroxi-L-prolina	11.5 (5 – 18)	11.8 (2,9 – 20,7)	13 (3 – 23)
46	2-Hidroxisovalerato	10.9 (5,6 – 181,8)	7.6 (2,6 – 12,6)	7,7 (0 – 19)
47	Propileno glicol	8,7 (4 – 88,9)	7.8 (2,1 – 10,5)	22,3 (19 – 25,6)
48	π -Metilhistidina	7.9 (1,9 – 13,9)	5.4 (2 – 8,8)	7,7 (5,8 – 9,6)
49	Uridina	5.2 (3,6 – 6,8)	5.3 (3 – 7,6)	3.1 (2,9 – 3,3)
50	2-Oxoisocaproato	5.0 (2,7 – 7,3)	5.4 (2,2 – 8,6)	28 (0 – 58)
51	Triptofano	4.5 (2,9 – 6,1)	4.2 (3,1 – 5,3)	54,5 (44,8 – 64,2)
52	τ -Metilhistidina	4.3 (2,9 – 5,7)	5.1 (1,4 – 8,8)	2,7 (0 – 5,9)
53	Inosina	3.9 (0,3 – 7,5)	9.7 (0 – 26,4)	0,05 (0,04 – 0,06)
54	Trigonelina	2.2 (0,9 – 3,5)	2.8 (0,7 – 4,9)	NE
55	1,5-anidro sorbitol	NQ	NQ	NQ
56	Acetona	NQ	NQ	NQ
57	Cafeína	NQ	NQ	NQ
58	Etanol	NQ	NQ	NQ
59	Glicerol	NQ	NQ	NQ
60	Ureia	NQ	NQ	NQ
61	DSS	NQ	NQ	NQ

Tabela 5: Médias e desvios padrão das concentrações metabólicas (μM) determinadas nas amostras do grupo controlo e nas amostras do grupo de doentes e retirados da base de dados HMDB para os metabolitos identificados. **NQ:** não quantificável; **NE:** não se apresenta descrito na literatura. A numeração dos metabolitos, presentes na tabela, corresponde à numeração que vai ser descrita ao longo deste capítulo e à numeração presente nas figuras A1 - A7 do anexo I.

4.4 Análise Univariada

Para a análise univariada foi realizado o teste de *t-Student* para um nível de confiança $p < 0,05$. Através deste teste realizado na plataforma MetaboAnalyst (Figura 19: Teste *t-Student*. As variáveis representadas acima da linha obtiveram um valor de p inferior a 0.05 e Tabela 6) verificou-se a existência de diferenças significativas entre o grupo controlo e o grupo de doentes para alguns aminoácidos (asparagina, aspartato, cisteína, metionina, ornitina, glutamato, serina e π -metilhistidina), glícidos (glucose e manose), ácidos orgânicos (lactato, citrato e 3-hidroxibutirato) e intermediários nucleosídeos (inosina e hipoxantina). Para as variáveis com um valor de p , calculado a partir do teste *t-Student*, inferior a 0.05, foram realizados gráficos de caixa de bigodes para se avaliar a existência de *outliers* que contribuam para a discriminação entre as amostras do grupo controlo e do grupo de doentes (anexo V).

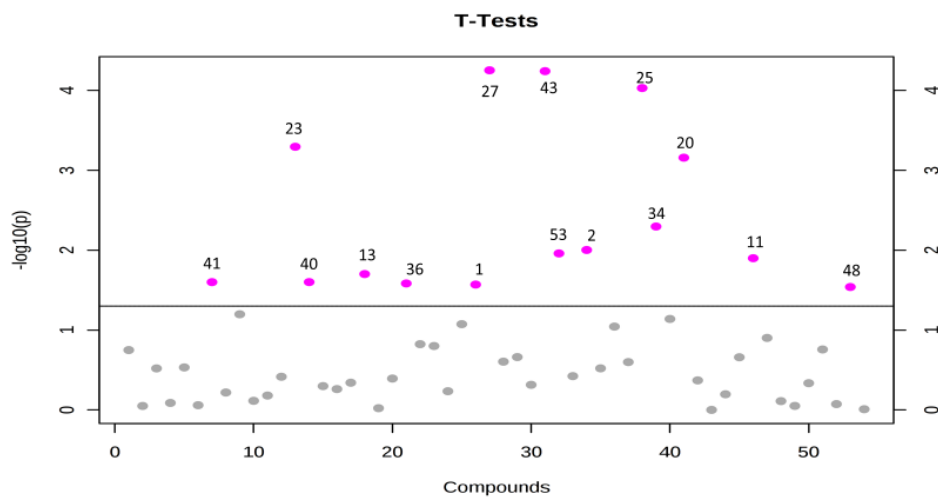


Figura 19: Teste *t-Sudent*. As variáveis representadas acima da linha obtiveram um valor de p inferior a 0.05

Tabela 6: Metabolitos e respectivos valores de p do teste *t-Sudent*

	Metabolito	Valor de p
27	Glutamato	5,59e-5
43	Hipoxantina	5,73e-5
25	Manose	9,33e-5
23	Asparagina	5,07e-4
20	Ornitina	6,94e-4
34	Metionina	0,005
2	Lactato	0,009
53	Inosina	0,011
11	Serina	0,013
13	Citrato	0,02
40	Aspartato	0,025
41	3-hidroxisobutirato	0,025
36	Cisteina	0,026
1	Glucose	0,027
48	π -metilhistidina	0,029

Através da Tabela 6, ordenada por valores crescentes de p é possível, em particular, verificar que o glutamato, a hipoxantina, a manose, a asparagina e a ornitina apresentaram um valor de $p < 0,005$. Na realidade através do anexo V verifica-se que estes metabolitos apresentam valores relativos às suas concentrações, muito diferentes no soro de doentes, quando comparados com indivíduos saudáveis.

4.5 Análise multivariada

Antes de se proceder à comparação entre amostras de indivíduos saudáveis e amostras de indivíduos com tuberculose, foram realizadas algumas análises preliminares. Primeiro foi realizada uma análise para selecionar a dimensão adequada de *bin* para proceder à análise dos dados através da abordagem *fingerprinting*. Posteriormente verificou-se a existência de *outliers* significativos no conjunto total das amostras. Após remoção de *outliers* significativos, foi avaliada a influência das variáveis externas (género, idade, índice de massa corporal (IMC) e hábitos tabágicos) na discriminação das amostras dentro de cada grupo. Além disso, foi realizada uma análise, para verificar se existiam diferenças significativas entre amostras de controlos, de doentes pulmonares e de doentes extrapulmonares. Estas duas últimas avaliações foram efetuadas, tanto para a análise do perfil metabólico como para a análise *fingerprinting*.

Após a realização destas análises preliminares, foi realizada a análise multivariada para todas as amostras, excluindo os *outliers*, tanto com recurso aos dados do perfil metabólico, como para os dados resultantes da análise *fingerprinting*.

4.5.1 Seleção da dimensão dos *bins*

O método *binning* escolhido para análise *fingerprinting* gera um grande número de variáveis. Na análise multivariada, o grande desequilíbrio entre o número de amostras e o número de variáveis gera resultados tendenciosos e pode gerar modelos pouco

robustos. De modo a evitar este tipo de situação foi otimizada a dimensão dos *bins* usados. Quanto maior o tamanho de bin utilizado, menor é o número de variáveis que irão ser utilizadas na análise multivariada. Foram testados tamanhos de *bins* de 0.01, 0.02, 0.03 e de 0.04 ppm. A utilização de *bins* de menor dimensão pode induzir desequilíbrio entre o número de variáveis e o número de amostras. Não foram selecionados *bins* de maior tamanho, pois assim, o número de sinais de metabolitos, que estes iriam incluir seria maior, dificultando a seleção das variáveis responsáveis pela discriminação dos grupos de amostras.

Primeiramente recorreu-se à construção de quatro modelos de PCA, um para cada conjunto de dados correspondentes a tamanhos diferentes dos *bins*. À exceção do modelo de PCA construído a partir de seis componentes principais, para o tamanho de bin de 0.01 ppm, todos os outros três modelos foram construídos a partir de cinco componentes principais. Observando a tabela 7, é possível verificar-se que o poder preditivo dos modelos (Q^2) aumentou com o aumento do tamanho dos *bins*. Além disso, a variação explicada pelos modelos (R^2X) também sofreu um ligeiro aumento. No entanto, como todos os valores de Q^2 são inferiores a 0,5, considera-se que nenhum dos modelos apresenta confiança estatística, não sendo possível retirar conclusões. Deste modo é necessário recorrer a modelos de PLS-DA para avaliar a influência do número de variáveis originais na discriminação dos grupos de amostras.

Tabela 7: Parâmetros relativos aos modelos de PCA usado para avaliar o tamanho dos *bins*

Tamanho <i>bin</i> (ppm)	R^2X	Q^2	Nº CPs	CP1 (%)	CP2 (%)
0.01	0.588	0.363	6	27,2	10,5
0.02	0.576	0.364	5	28,6	11,2
0.03	0.592	0.366	5	29,4	11,4
0.04	0.605	0.375	5	30,1	12,1

Observando os gráficos de *scores* para os quatro modelos de PLS-DA e os respetivos testes de permutação (Figura D do anexo VI), não é possível observar alterações significativas entre a distribuição das amostras associada ao número de variáveis originais utilizado. No entanto, olhando para os parâmetros da validação descritos na tabela 8, para os quatro modelos, é possível verificar algumas diferenças. Os valores de R^2 e de Q^2 do teste de permutação tendem a diminuir com a diminuição do

número de variáveis originais, apresentando Q^2 valores sempre adequados. Apesar do valor de R^2 do teste de permutação, na análise realizada com bins de tamanho igual a 0.04 ppm, ser superior a 0,4, este deve ser aceite. A tendência é do valor de R^2 do teste de permutação diminuir, ou com o aumento do número de amostras do grupo de doentes, ou com a diminuição do número de variáveis originais. Simultaneamente, o modelo gerado para bins de tamanho igual a 0.04 ppm apresenta uma maior percentagem de variação explicada (R^2X 0,521). Apesar do poder preditivo dos modelos de PLS-DA (Q^2) diminuir com a redução do número de variáveis originais, apresenta para bins de 0.04 ppm um valor aceitável (próximo de 0,5).

Tabela 8: Parâmetros relativos aos modelos de PLS-DA usado para avaliar o tamanho dos bins

	R^2X	R^2Y	Q^2	Nº CPs	CP1 (%)	CP2 (%)	p ANOVA	cv.	R^2 perm	Q^2 perm
0.01 ppm	0.466	0.863	0.574	4	17,2	17,5	$6.16e^{-12}$		0.639	-0.151
0.02 ppm	0.491	0.826	0.515	4	18,6	17,5	$1.29e^{-10}$		0.561	-0.239
0.03 ppm	0.505	0.797	0.435	4	20	17,1	$4.29e^{-8}$		0.508	-0.276
0.04 ppm	0.521	0.788	0.471	4	20,8	17	$1.04e^{-9}$		0.457	-0.295

4.5.2 Remoção de outliers

Um modelo de PCA foi construído a partir dos dados, para a análise do perfil metabólico e outro para a análise *fingerprinting*. Observando os gráficos de *scores* dos modelos de PCA para a análise do perfil metabólico (Figura 20) e para a análise *fingerprinting* (Figura 21) foi possível observar a presença de *outliers* no conjunto total de amostras. No caso da análise do perfil metabólico é possível constatar que as amostras C33 e D20, estão localizadas, a uma distância considerável, isto é, fora do limite do *Hotelling T²*. Este define a área normal correspondente a 95% de confiança, cuja a extensão é limitada pela elipse representada (L. Eriksson 2006). No caso da análise *fingerprinting* observa-se que a amostra C42 é o *outlier* mais evidente. Além deste *outlier*, na análise *fingerprinting*, observou-se que as amostras C33, D20 e D12 se apresentavam como *outliers*. Verificou-se que os sinais dos metabolitos do espectro da amostra C42

apresentavam-se um pouco deformados, devido ao mau *shimming* apresentado pelo espectro. Esta amostra não se apresentou como *outlier* na análise do perfil metabólico pois, com a ajuda do Chenomx, foi possível contornar este problema, fazendo uma quantificação e identificação razoáveis dos metabolitos. A amostra C33 e D20 apresentaram-se como *outliers* para ambas as análises. No espectro da amostra C33, alguns sinais de metabolitos (ácidos orgânicos) apresentavam-se bastante intensos. Este facto pode ser consequência do indivíduo ao qual pertence a amostra C33 apresentar um IMC superior a quarenta. Já o espectro da amostra D20, apresenta todos os sinais com intensidade elevada, relativamente a outros espectros. Isto pode ter ocorrido devido a erros durante a manipulação das amostras. A amostra D12, apesar de ser um *outlier*, não foi excluída da análise estatística. O número de amostras do grupo de doentes é relativamente baixo, quando comparado com o número de amostras do grupo controlo. Assim, não é muito aconselhável remover amostras do grupo de doentes, a não ser que exista uma razão concreta para o efetuar.

Na sequência desta avaliação, as amostras C33 e D20 foram retiradas para ambas as análises. A amostra C42 apenas foi retirada para a análise realizada com os dados da análise *fingerprinting*.

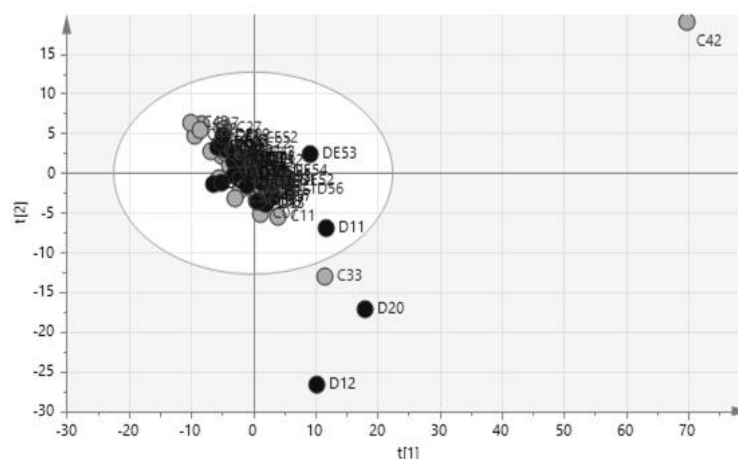


Figura 20: Gráfico de *scores* do modelo de PCA utilizado para a exclusão de *outliers*, o caso da análise *fingerprinting*

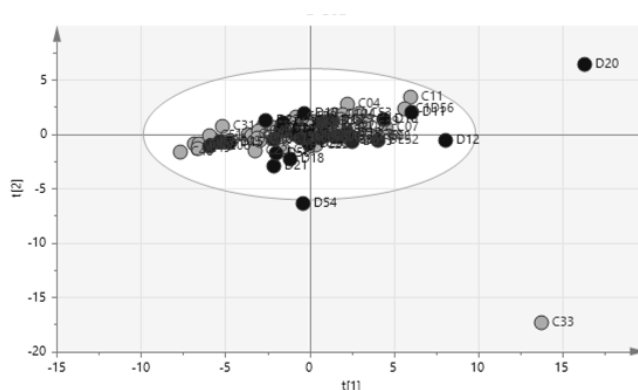


Figura 21: Gráfico de *scores* do modelo de PCA utilizado para a exclusão de *outliers*, o caso da análise do perfil metabólico

4.5.3 Análise de variáveis externas

A composição metabólica do soro sanguíneo é considerada relativamente estável, tendo em conta o papel homeostático do sangue no corpo humano. Contudo, diversas variáveis externas (ex: dieta, estilo de vida, idade, género) podem ser responsáveis por alguma da variabilidade entre indivíduos, influenciando a interpretação dos efeitos relacionados coma doença. Por esse motivo, deve avaliar-se o efeito das variáveis externas em cada um dos grupos de amostras. Neste trabalho foi investigada a influência do género, da idade, do IMC e dos hábitos tabágicos dos indivíduos, no perfil metabólico do soro, tanto para os indivíduos saudáveis, como para os indivíduos com tuberculose.

4.5.3.1 Análise de variáveis externas em indivíduos saudáveis

Olhando para as tabelas 9 e 10, é possível verificar que os modelos de PLS-DA (figuras E1.1 – E1.4 do anexo VII) construídos para avaliar a influência dos hábitos tabágicos na composição metabólica do soro, não são de todo preditivos, apresentando valores de Q^2 negativos para ambas as análises. O mesmo se verifica para a variável idade, no caso da análise do perfil metabólico. Observando os gráficos de *scores* dos modelos de PLS-DA, para cada variável externa, de cada análise, apenas é visível a formação de agrupamentos de amostras para as variáveis gênero e IMC. Esta separação das amostras, no entanto, não pode ser levada em conta, dado os valores de Q^2 serem inferiores a 0,5. Os valores de p dos modelos para as variáveis gênero e idade e para as variáveis gênero e IMC, são inferiores a 0.05, no caso da análise *fingerprinting* e da análise do perfil metabólico, respetivamente. No entanto, os modelos de PLS-DA construídos a partir dos dados da análise do perfil metabólico e de *fingerprinting*, os valores de Q^2 para todos os modelos são inferiores a 0,5. Na globalidade, os parâmetros de validação para os modelos não permitem ter confiança estatística para aceitá-los como verdadeiros. Assim, é rejeitada a hipótese das variáveis externas afetarem significativamente a composição metabólica do soro sanguíneo das amostras controlo.

Tabela 9: Parâmetros relativos aos modelos de PLS-DA usados para avaliar as variáveis externas nas amostras controlo para a análise do perfil metabólico

	R^2X	R^2Y	Q^2	Nº CPs	CP1 (%)	CP2 (%)	p ANOVA	CV-	R^2 perm	Q^2 perm
Gênero	0,349	0,651	0,383	2	29	5,88	$1,52 \cdot 10^{-4}$		0,367	-0,114
Idade	0,331	0,376	-0,082	2	27,7	5,45	1		0,083	-0,099
IMC	0,359	0,355	0,108	2	28,6	7,33	0,009		0,350	-0,080
Hábitos Tabágicos	0,358	0,287	-0,128	2	7,03	28,8	1		0,341	-0,064

Tabela 10: Parâmetros relativos aos modelos de PLS-DA usados para avaliar as variáveis externas nas amostras controle para a análise *fingerprinting*

	R ² X	R ² Y	Q ²	Nº CPs	CP1 (%)	CP2 (%)	p ANOVA	cv-	R ² perm	Q ² perm
Género	0.377	0.656	0.363	2	30,3	7,4	3.97*10 ⁴		0.353	-0.127
Idade	0.429	0.272	0.13	2	26,7	16,2	0.0048		0.272	-0.141
IMC	0.484	0.401	0.075	3	29	13	0.99		0.457	-0.188
Hábitos Tabágicos	0.427	0.185	-0.156	2	18,3	24,4	1		0.335	-0.063

Segundo o gráfico de *loadings*, os principais metabolitos responsáveis pela visível separação entre homens e mulheres são o glutamato, a prolina, a carnitina e o citrato. Destes quatro metabolitos apenas a carnitina foi descrita como estando em maiores concentrações em indivíduos do género masculino (Mittelstrass et al. 2011). No caso do agrupamento de amostras observado para indivíduos com um IMC elevado, este é causado por metabolitos como o 2-oxoisocaproato, o 2-hidroxibutirato, a leucina e a valina. Destes quatro metabolitos apenas a valina e a leucina foram descritos com estando relacionados com o índice de massa corporal dos indivíduos (Moore et al. 2014).

4.5.3.2 Análise de variáveis externas em indivíduos com tuberculose

Olhando para as tabelas 11 e 12, é possível verificar que os modelos de PLS-DA (Figuras F1-F4 do anexo VIII) construídos para avaliar a influência dos hábitos tabágicos na composição metabólica do soro, não são de todo preditivos, apresentando valores de Q² negativos para ambas as análises. Observando os gráficos de *scores* dos modelos de PLS-DA, para cada variável externa, de cada análise, apenas é visível a formação de agrupamentos de amostras para as variáveis género. Esta separação das amostras, no entanto, não pode ser levada em conta, dados valores de Q² inferiores a 0,5. Os valores de *p* dos modelos para todas as variáveis externas, no caso da análise *fingerprinting* e da análise do perfil metabólico, são superiores a 0.05. No entanto, os modelos de PLS-DA construídos a partir dos dados da análise do perfil metabólico e de *fingerprinting*, os

valores de Q^2 para todos os modelos são inferiores a 0,5. Na globalidade, os parâmetros de validação para os modelos não permitem ter confiança estatística para aceitá-los como verdadeiros. Assim, é rejeitada a hipótese de as variáveis externas afetarem significativamente a composição metabólica do soro sanguíneo das amostras controle.

Tabela 11: Parâmetros relativos aos modelos de PLS-DA usados para avaliar as variáveis externas nas amostras de doentes para a análise do perfil metabólico

	R^2X	R^2Y	Q^2	Nº CPs	CP1 (%)	CP2 (%)	p ANOVA	cv.	R^2 perm	Q^2 perm
Género	0,4	0,807	0,092	3	18	13	0,88		0,737	-0,205
Idade	0,402	0,621	0,233	3	13,5	15,7	0,21		0,563	-0,184
IMC	0,307	0,413	0,013	2	22,2	8,44	0,85		0,395	-0,099
Hábitos Tabágicos	0,355	0,416	-0,21	2	25,4	10	1		0,434	-0,045

Tabela 12: Parâmetros relativos aos modelos de PLS-DA usados para avaliar as variáveis externas nas amostras de doentes para a análise *fingerprinting*

	R^2X	R^2Y	Q^2	Nº CPs	CP1 (%)	CP2 (%)	p ANOVA	cv.	R^2 perm	Q^2 perm
Género	0.58	0.855	0.351	4	21,1	21,9	0.15		0.808	-0.122
Idade	0.534	0.491	0.15	3	27,7	16,3	0.59		0.489	-0.138
IMC	0.44	0.327	0.055	2	32,5	11,5	0.57		0.278	-0.089
Hábitos Tabágicos	0.462	0.346	-0.102	2	26,2	20	0.99		0.413	-0.019

Segundo os gráficos de loadings os principais metabolitos responsáveis pela visível separação entre homens e mulheres são a isoleucina e o citrato. Apenas o citrato foi observado tanto para o grupo controle, como para o grupo de doentes, como sendo um dos metabolitos responsáveis pela separação entre indivíduos do género feminino e masculino.

4.5.4 Discriminação entre amostras de indivíduos saudáveis e de indivíduos com tuberculose

O sangue é o principal fluido transportador de gases, nutrientes, hormonas e de metabolitos no corpo humano, mediando as interações metabólicas entre diferentes órgãos/tecidos. Por esta razão, é expectável que a composição metabólica do soro sanguíneo seja um reflexo do fenótipo de doenças, com a tuberculose.

Após a realização das análises preliminares, procedeu-se á análise multivariada dos dados para os dois grupos experimentais. Neste passo visa-se avaliar a discriminação entre grupos de amostras, tentando encontrar as principais variáveis que mais influenciam a separação entre as amostras de indivíduos saudáveis e as amostras de indivíduos com tuberculose. Para isso recorreu-se a dois métodos metabolómicos para tornar os resultados obtidos mais robustos.

Como os pacientes incluídos neste estudo apresentaram dois tipos de tuberculose, foi necessário avaliar o impacto causado por esta heterogeneidade na composição metabólica do soro sanguíneo. Deste modo, recorreu-se à construção de modelos para o método *fingerprinting* e para o perfil metabólico, para tentar discriminar indivíduos com tuberculose pulmonar de indivíduos com tuberculose extrapulmonar.

Como se observa na tabela 13, ambos os valores de Q^2 dos modelos de PCA para os dois métodos, são inferiores a 0,5. Assim, como não existe confiança estatística, não é possível retirar conclusões a partir dos modelos de PCA, sendo necessário a utilização do PLS-DA. Os modelos de PLS-DA de ambos os métodos apresentam parâmetros de validação pobres ($Q^2 < 0,5$; $p_{CV-ANOVA} > 0,05$; $R^2_{perm} > 0,4$). Assim, dado que não existe confiança estatística para retirar conclusões a partir dos modelos de PLS-DA, rejeita-se o facto de o tipo de tuberculose influenciar na separação das amostras do grupo de doentes.

Tabela 13: Parâmetros de validação dos modelos da análise multivariada para a comparação de amostras do grupo de doentes pulmonares e de doentes extrapulmonares

		R^2X	R^2Y	Q^2	Nº CPs	CP1 (%)	CP2 (%)	p ANOVA	cv-	R^2 perm	Q^2 perm
Perfil metabólico	PCA	0.356	-----	0.12	2	23,8	11,8	-----	-----	-----	-----
	PLS-DA	0.383	0,833	0.311	3	10,7	16,1	0.27	-----	0.716	-0,188
Fingerprinting	PCA	0.48	-----	0.226	2	33,8	14,2	-----	-----	-----	-----
	PLS-DA	0.208	0,685	0,0212	2	12,1	8,62	0.37	-----	0.471	-0.061

Como não foi possível verificar diferenças entre amostras de indivíduos com tuberculose pulmonar e de indivíduos com tuberculose extrapulmonar, estas foram colocadas dentro de um mesmo grupo (grupo de doentes). Assim, apenas foi realizada a análise discriminatória entre indivíduos com tuberculose e indivíduos saudáveis

A comparação de dados de doentes com controlos, foi efetuada inicialmente por PCA, construindo dois modelos (Figura 22), um para cada tipo de análise metabolômica. Relativamente á percentagem de variação dos dados, explicada por ambos os modelos de PCA (Tabela 14), é possível verificar que o modelo construído a partir do método *fingerprinting* explica uma maior variação (R^2X 60,5%) dos dados, quando comparado com o modelo construído a partir da análise do perfil metabólico (R^2X 32,3%). Os valores preditivos para ambos os modelos foram inferiores a 0,5 (Q^2_{pm} 0,172; Q^2_{fing} 0,375). Através dos gráficos de scores apresentados na figura 18 não existindo confiança estatística para concluir sobre a discriminação entre os 2 grupos de amostras. Assim, é necessário a construção de modelos de PLS-DA.

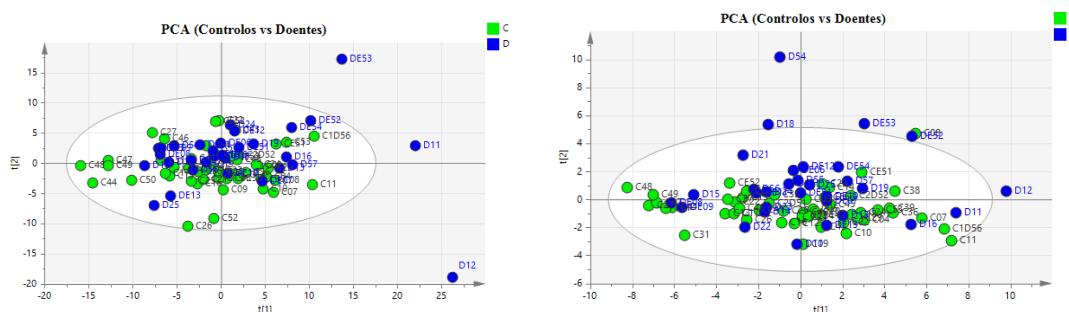


Figura 22: Gráficos de scores do modelo de PCA referentes à análise fingerprinting (imagem da esquerda) e à análise do perfil metabólico (imagem da direita), usado para comparar amostras do grupo controlo e do grupo de doentes.

Tabela 14: Parâmetros de validação dos modelos de PCA

	R ² X	Q ²	Nº CPs	CP1 (%)	CP2 (%)
Perfil metabólico	0.323	0.172	2	24,6	7,77
Fingerprinting	0.605	0.375	5	30,1	12,1

No modelo de PLS-DA, para a análise *fingerprinting*, foi possível observar, através do gráfico de *scores* (Figura 23 e tabela 15), que as amostras de indivíduos saudáveis e de indivíduos com tuberculose, tendem a agrupar-se em grupos distintos. Este modelo, construído com quatro variáveis latentes, apresenta R²X, R²Y e Q² de 0,519, 0,791 e 0,478, respetivamente. Além disso, o valor de *p* do teste ANOVA da validação cruzada é de 5.75e⁻¹⁰. Pelo que é demonstrado pelo teste de permutações, é possível observar-se um R² e um Q² de 0.463 e -0.293, respetivamente (Figura G 1 do anexo IX). Apesar de o R² ser um pouco elevado, no teste de permutação, este segue uma tendência em diminuir o seu valor com o aumento do número de amostras.

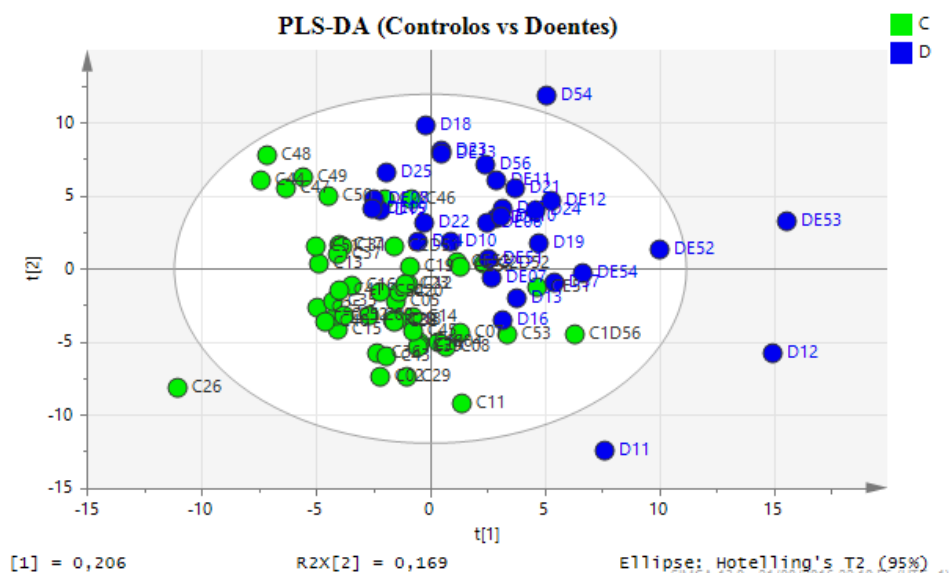


Figura 23: Gráfico de scores do modelo de PLS-DA para a análise *fingerprinting*

Observando o gráfico de scores do modelo de PLS-DA, realizado para a análise do perfil metabólico (Figura 24 e tabela 15), podemos verificar uma clara separação entre amostras do grupo controlo e do grupo de doentes. Este modelo, construído com 3 variáveis latentes, apresenta um R^2X de 0,352, um R^2Y de 0,773 e um Q^2 de 0,546. Além disso, o valor de p do teste ANOVA da validação cruzada é de $1,15e^{-11}$.

Pelo que é demonstrado pelo teste de permutações (Figura G 2 do anexo IX), é possível observar-se um R^2 e um Q^2 menor que 0,4 e 0,05, respetivamente. Os parâmetros da validação cruzada indicam a boa qualidade do modelo, dando-nos confiança estatística para o aceitar. A partir deste modelo é possível concluir que existe uma separação entre as amostras do grupo controlo e as amostras do grupo de doentes

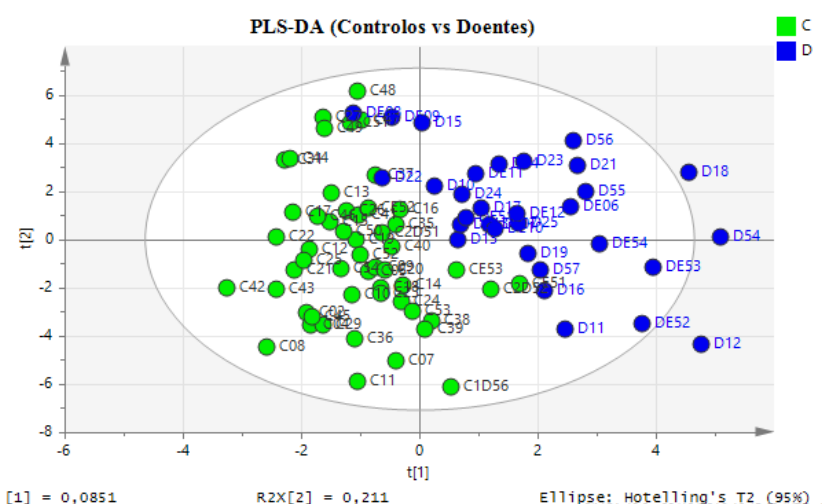
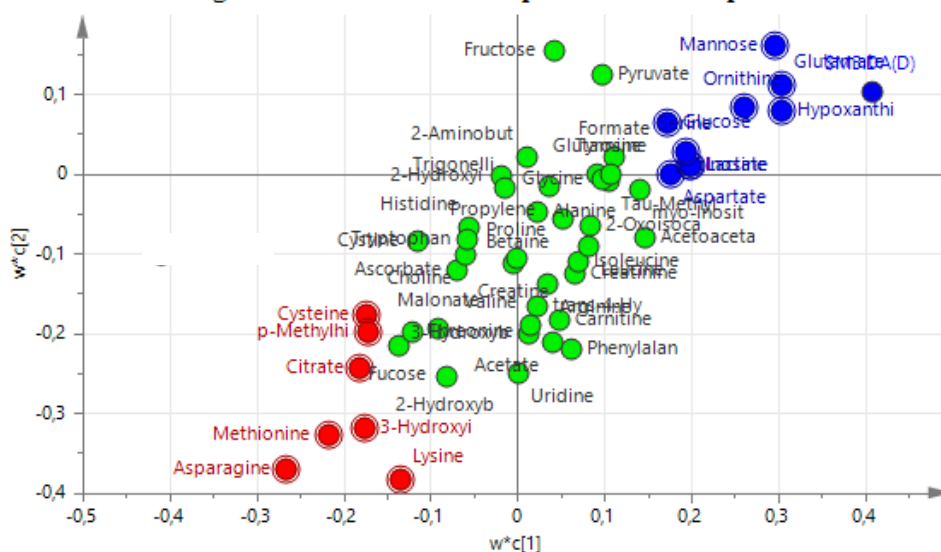


Figura 24: Gráfico de *scores* do modelo de PLS-DA para a análise do perfil metabólico

Tabela 15: Parâmetros de validação dos modelos de PLS-DA

	R^2X	R^2Y	Q^2	Nº CPs	CP1 (%)	CP2 (%)	p ANOVA	cv- ANOVA	R^2 perm	Q^2 perm
Perfil metabólico	0,352	0,773	0,546	3	8,51	21,1	$1,15e^{-11}$		0,369	-0,273
Fingerprinting	0,519	0,791	0,478	2	20,6	16,9	$5,75e^{-10}$		0,463	-0,293

Gráfico de loadings do modelo de PLS-DA para a análise do perfil metabólico



Para a análise *fingerprinting*, o gráfico de *loadings* (Figura 26) apresenta uma maior complexidade, causada pelo elevado numero de variáveis. Assim foi necessário

usar os VIPs maiores que 1 (Figura G 3 do anexo IX) para se proceder a uma melhor seleção das variáveis responsáveis pela separação das amostras observada na figura 26.

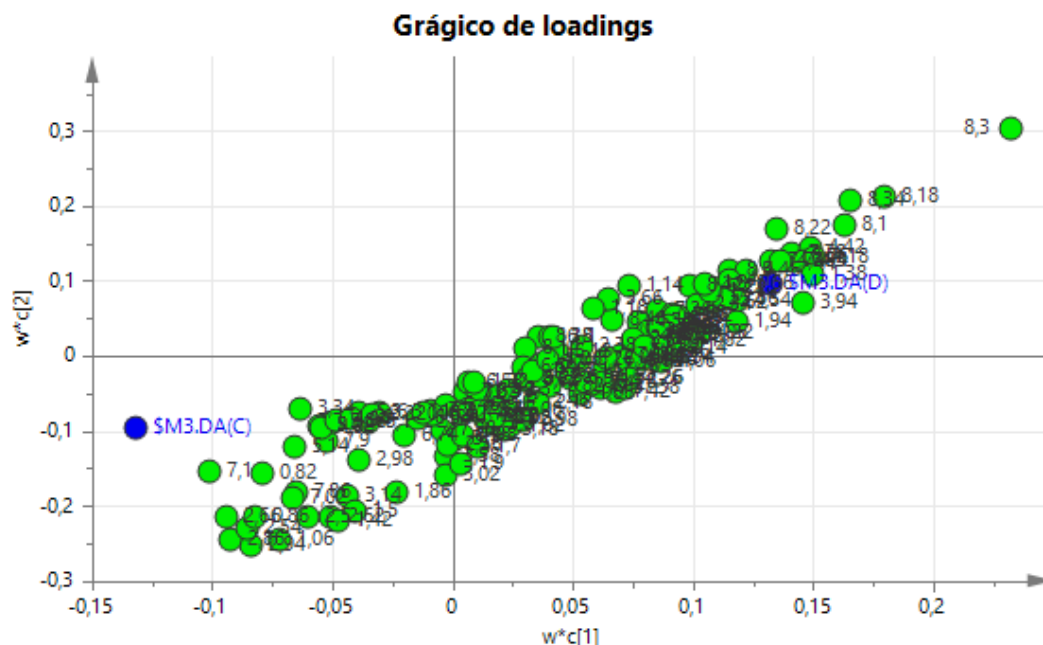


Figura 26: Gráfico de *loadings* do modelo de PLS-DA para a análise *fingerprinting*. Os pontos a azul representam a tendência do conjunto de amostras de cada grupo

Após a seleção dos principais bins responsáveis pela separação das amostras, e após comparação com os resultados obtidos pela análise do perfil metabólico foi possível chegar aos resultados descritos na tabela 16.

Alguns *bins* considerados importantes na separação das amostras, contêm sinais que não puderam ser identificados (tabela 16). Os *bins* onde se encontravam os sinais da histidina (7.1 ppm e 7.86 ppm), foram considerados relevantes para a separação das amostras. Contudo, devido à malformação do sinal da histidina em alguns espectros, esta foi excluída da posterior análise de enriquecimento do conjunto de metabolitos.

Tabela 16: Metabolitos e respectivos sinais responsáveis pela separação entre as amostras do grupo controle e as amostras do grupo de doentes utilizando a análise do perfil metabólico e a análise *fingerprinting*

Metabolitos	Bins (ppm)
Hipoxantina	8.22; 8.18
Inosina	8.34; 8.22
Lactato	4.18; 1.38
Manose	3.94; 3.86
Serina	3.94
Ornitina	1.94; 3.06
Glucose	4.62; 3.82
Glutamato	2.34
Aspartato	3.9
Citrato	2.66; 2.54; 2.5
Asparagina	2.94; 2.86
3-hidroxisobutirato	1.06
II-metilhistidina	7.02
Metionina	2.62
Lisina	1.86; 1.5; 1.42
Cisteína	3.14
Não identificados	8.3; 8.1; 7.94; 7.78; 7.74; 6.66; 4.54; 4.46; 4.38; 4.3; 4.22; 2.98; 0.86

4.6 Análise das vias metabólicas

Através da análise uni e multivariada foi possível identificar quinze metabolitos comuns. Observa-se, no entanto, que a lisina foi considerada como um importante metabolito na discriminação dos grupos de amostras, apenas na análise multivariada. Assim, verifica-se que, com três análises diferentes, se obtiveram resultados similares. A lisina, apesar de não ter sido identificada como significativa, através da análise univariada, foi utilizada para a análise de enriquecimento do conjunto de metabolitos. Isto porque, a análise multivariada, apresenta-se como um método mais robusto, comparativamente à análise univariada.

A análise metabólica é uma ferramenta importante para a compreensão das alterações metabólicas causadas pela tuberculose. Ao analisarmos os metabolitos, significativamente alterados pela infecção de *Mtb*, é possível verificar quais as vias metabólicas alteradas em consequência da doença.

Através da análise de enriquecimento do conjunto de metabolitos (MSEA) foi possível verificar que as vias metabólicas mais afetadas pela infecção de *Mtb* correspondiam à biossíntese de proteína, à reciclagem de amônia, ao metabolismo do aspartato e ao metabolismo da metionina (Figura 27 e Tabela 17). Todos os 4 processos metabólicos apresentaram valores de *p* inferiores a 0.05 e uma taxa de descoberta de falsos positivos (FDR) inferior a 50% (Tabela 17)

Tabela 17:. Principais processos metabólicos alterados, como respetivo número de metabolitos, valores de *p* e taxa de descoberta de falsos positivos (FDR)

Conjunto de metabolitos	Nº de metabolitos	Valor de <i>p</i>	FDR (%)
Biossíntese de proteína	5	1,38e ⁻⁵	0,1
Reciclagem de amoníaco	3	0,0041	16,4
Metabolismo da metionina	3	0,0095	25,2
Metabolismo do aspartato	2	0,0208	41,6

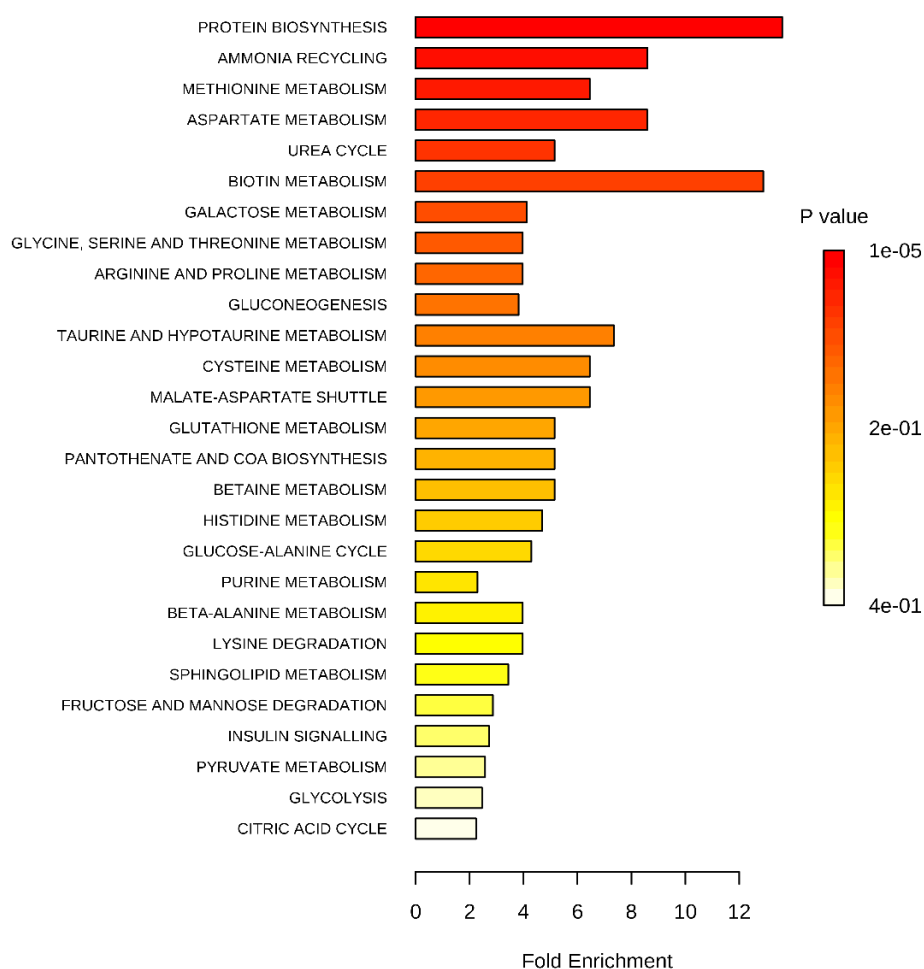


Figura 27: Análise de enriquecimento do conjunto de metabolitos representativa dos processos metabólicos principais alterados no soro de indivíduos com tuberculose ativa.

As concentrações da hipoxantina e da inosina foram encontradas aumentadas no soro de doentes, comparativamente com os controlos. Estes dois metabolitos estão associados à biossíntese de nucleótidos. Tal como foi demonstrado pelo trabalho de Zhou (Zhou et al. 2013), o aumento da biossíntese de nucleótidos está associado à divisão ativa das células inflamatórias hospedeiras. Um estudo realizado por Shin (Shin et al. 2011) demonstrou que o aumento da biossíntese de nucleótidos pode estar também associado à utilização de precursores de nucleótidos do hospedeiro, pelo microrganismo, para que este se possa multiplicar.

O lactato encontrou-se aumentado no soro de doentes. O lactato resulta da glicólise anaeróbia, que, neste caso, se encontra aumentada, devido à infeção por *Mtb*. Como já foi descrito em estudos anteriores, a infeção granulomatosa causada pela infeção, pode gerar a hipoxia dos tecidos (Via et al. 2008). Isto é, pode levar à escassez de oxigénio nos tecidos. Assim, as células presentes nos tecidos infetados recorrem à glicólise anaeróbia para produzir energia.

Foi demonstrado num estudo que a atividade da metionina adenosiltransferase (MAT) se encontra reduzida, devido à hipoxia (Avila et al. 1998). Esta enzima é responsável pela conversão da metionina em adenosilmetionina, que por sua vez vai ser convertida em metionina ou originar cisteína (Finkelstein 1990). Dado a reduzida atividade da MAT em condições de hipoxia, os níveis de cisteína e de metionina irão ser reduzidos. A serina é utilizada para formar cisteína. Uma vez que o metabolismo da metionina se encontra diminuído, a serina não vai ser utilizada, sendo isto uma das causas do seu aumento no soro.

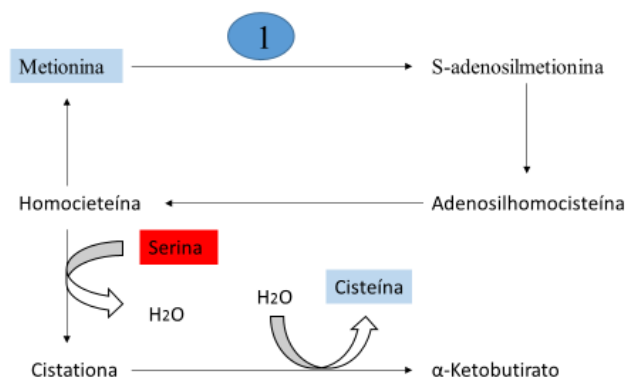


Figura 28: Metabolismo da metionina **1**: metionina adenosiltransferase. A vermelho e a azul estão marcados os metabolitos com níveis elevados e reduzidos, respetivamente no soro de doentes

A asparagina está presente em concentrações reduzidas no soro de doentes. A asparagina atravessa a membrana do macrófago por um sistema de transporte que até à data é desconhecido. Esta vai ser utilizada por *Mtb* como fonte de nitrogénio (Bottai et al. 2014). Bottai e colaboradores demonstrou ainda que a asparagina se acumula nos vacúolos micobacterianos dentro dos macrófagos infetados, sendo que esta vai ser convertida em aspartato e em amónia no interior dos macrófagos infetados. O aspartato é uma fonte importante de nitrogénio usada por *Mtb* durante a infeção. Este mecanismo vai permitir à bactéria resistir ao ambiente ácido dos fagossomas e sobreviver dentro das células (Bottai et al. 2014).

A hidrólise da arginina pela arginase-1 origina ornitina e ureia. Em hipoxia, a arginase-1 encontra-se aumentada, o que vai induzir a supressão das células T (Duquecorrea et al. 2014). Consequentemente existe um aumento da quantidade de ornitina sintetizada por esta via metabólica. A ornitina também é produzida através do ciclo da ureia. *Mtb* utiliza os aminoácidos do hospedeiro como fonte de nitrogénio, libertando posteriormente amónia para o citosol dos macrófagos (Bottai et al. 2014). Assim é necessário transformar a amónia em ureia através do ciclo da ureia, sendo produzida mais ornitina (Voet & Voet 1996).

A concentração diminuída de aminoácidos no soro de doentes observada neste trabalho poderá estar relacionada com o facto de estes serem convertidos em intermediários do ciclo do citrato. Foi descrito anteriormente que a diminuição de aminoácidos como a leucina, a isoleucina, a valina e a metionina poderá representar uma diminuição dos intermediários correspondentes do ciclo do citrato, devido à provação de energia (Chen et al. 2013). Foi observado neste trabalho uma diminuição do citrato em indivíduos com tuberculose, o que coincide com a afirmação anterior.

A proteólise ocorrida durante a lise necrótica dos macrófagos e das células T poderá estar associada com o aumento dos níveis de glutamato e de aspartato observado neste estudo. Este aumento foi observado por Somashekar e colaboradores (Bagganahalli S. Somashekar et al. 2012).

Apesar de se observar um aumento dos níveis do glutamato, da ornitina, da serina e do aspartato observou-se também a diminuição dos níveis da cisteína, da asparagina, da lisina e da metionina, no soro de doentes. A diminuição dos níveis de alguns aminoácidos pode também estar associada à sua utilização pelas células hospedeiras para produzir proteínas responsáveis pelo combate contra *M. tuberculosis* e para a sua replicação.

Dado que foram encontrados aminoácidos com concentrações reduzidas no soro de doentes, estes podem estar a ser utilizados para a gliconeogénese (Voet & Voet 1996). Este facto é suportado pelos níveis de glucose aumentados no soro de doentes. Os baixos níveis de π -metilhistidina encontrados no soro de indivíduos com tuberculose, contrariam resultados demonstrados num estudo de Zhou e colaboradores (Zhou et al. 2013). O 3-hidroxisobutirato foi encontrado a concentrações reduzidas no soro de doentes, não se tendo encontrado justificação para tal.

A manose foi encontrada elevada no soro de pacientes. Este resultado é consistente com estudos anteriores (Frediani et al. 2014; Weiner et al. 2012) onde é proposto que este aumento deriva da degradação da parede celular das micobactérias. Estes estudos revelam que a manose é um importante constituinte da parede celular da *Mtb*, fazendo parte dos lipoarabinomananos (LAMs).

Capítulo V: Conclusões

Neste trabalho, a assinatura metabólica da tuberculose foi investigada através de uma abordagem metabolômica envolvendo a análise do soro de indivíduos com tuberculose comparado com o soro de indivíduos saudáveis.

Com base na hipótese de que a tuberculose talvez possa ter impacto no metabolismo do soro sanguíneo, foram analisadas 54 amostras de indivíduos saudáveis e 32 amostras de indivíduos com tuberculose (incluindo indivíduos com tuberculose pulmonar e extrapulmonar), com recurso à espectroscopia de ressonância magnética nuclear de próton (^1H NMR).

Apesar do NMR constituir uma boa metodologia para a quantificação do metaboloma de biofluidos, este não consegue detetar o metaboloma completo, maioritariamente devido à sua reduzida sensibilidade e ao tratamento da amostra. Assim, é sempre necessário otimizar os procedimentos envolvidos na preparação de amostras e na aquisição de espectros. Apesar do processo de ultrafiltração ser um ótimo método para extração dos metabolitos, este também precisa de ser otimizado. Neste trabalho pode-se concluir que é necessário um grande número de lavagens dos filtros para remover o glicerol na sua totalidade. Outra conclusão retirada a partir da otimização do protocolo, diz respeito ao local onde os filtros são deixados com água *overnight*, após lavagem. Foi verificado que guardar os filtros em locais onde estejam presentes solventes orgânicos, como o etanol e a acetona, é suficiente para a sua deteção nos espectros das amostras.

Mesmo com uma sequência de pulso NOESY diferente da recomendada pelo Chenomx, foi conseguida uma boa quantificação e identificação dos metabolitos neste software. Verificou-se também, que apesar do grande número de metabolitos presentes nas bases de dados do Chenomx, não foi possível identificar todos os sinais presentes nos espectros.

A análise multivariada dos espectros de soro sanguíneo permitiu a construção de modelos de classificação robustos, discriminando indivíduos saudáveis e indivíduos com tuberculose. Os parâmetros obtidos permitiram validar os modelos de PLS-DA, tanto para a análise do perfil metabólico, como para a análise *fingerprinting*. Através das variáveis discriminativas no soro de doentes, foi revelado que os níveis de intermediários de nucleosídeos (hipoxantina e inosina), de glícidos (manose e glucose), de aminoácidos (ornitina, glutamato, serina e aspartato) e de ácidos orgânicos (lactato) se encontravam

aumentados. Já aminoácidos como a asparagina, a metionina, a cisteína, a lisina e a π -metilhistidina, e ácidos orgânicos, como o 3-hidroxisobutirato e o citrato, foram encontrados em menores concentrações no soro dos indivíduos com tuberculose.

Assim, foi possível verificar que através de dois métodos diferentes se obtiveram resultados similares, dando-lhes mais credibilidade.

Através da análise univariada foi possível identificar catorze metabolitos responsáveis pela separação das amostras. Apenas a lisina e o glutamato, não foram identificados pela análise univariada como existindo em concentrações distintas. Assim, podemos concluir que o uso da análise univariada, com recurso ao MetaboAnalyst 3.0, nos dá um resultado bastante similar ao obtido pela análise multivariada.

Conclui-se que vários processos metabólicos estão alterados nos indivíduos com tuberculose. Contudo, estas alterações não são específicas da tuberculose, podendo ser encontradas em outras patologias. Assim é necessário recorrer à utilização de outras ómicas para fazer uma relação com os resultados obtidos neste e noutros estudos relacionados com a tuberculose, de maneira a descobrir um conjunto de biomarcadores específicos para a tuberculose. Para estender o número de metabolitos quantificados pode-se recorrer à espectrometria de massa.

Muitos foram os indivíduos que tinham começado o tratamento contra a tuberculose, na altura da recolha do sangue. Este problema não foi possível ser contornado dados os protocolos do Serviço Nacional de Saúde no acompanhamento deste tipo de doentes. Dado que a taxa de incidência da tuberculose em Portugal, é muito baixa, comparativamente com outros países, acresceu a dificuldade de incluir no estudo indivíduos com tuberculose sem co-morbilidades associadas.

Capítulo VI: Perspetivas futuras

Este trabalho decorreu no âmbito do projeto “TBomics - an OMICS approach for diagnosing tuberculosis”, New INDIGO Partnership Programme on Biotechnology applied to Human Health (INDIGO-DBT2-062), ERA-NET Project, 2014-2016 com as autorizações da Comissão de Ética da ARS de Lisboa e Vale do Tejo (Parecer 057/CES/INV/2014) e da Comissão Nacional de proteção de dados (Autorização 5985/2014). O projeto tem como objetivo identificar biomarcadores específicos para a tuberculose ativa, latente e extrapulmonar em soros de pacientes, usando a espectroscopia de ressonância magnética nuclear e a espectrometria de massa. Para atingir a sensibilidade e especificidade adequadas, a abordagem utilizada consiste na combinação de resultados de metabolómica e proteómica.

Para este trabalho, inicialmente estava prevista a utilização de amostras de pacientes com tuberculose latente. Dado a dificuldade em encontrar indivíduos com esta patologia no estado de latência, não foi conseguido um número suficiente de amostras até à data. Assim é necessário a obtenção de mais amostras quer de paciente com tuberculose pulmonar e extrapulmonar ativa, quer de pacientes com tuberculose latente. Além disso é preciso utilizar a abordagem proteómica para adicionar informação aos resultados obtidos até à data.

Os resultados obtidos neste trabalho vão ser posteriormente comparados com os resultados obtidos pela equipa indiana, envolvida no projeto. Para além disso é necessário recorrer-se á validação dos biomarcadores encontrados, para posterior aplicação no diagnóstico da tuberculose.

Capítulo VII: Bibliografía

- Alonso, A., Marsal, S. & Julià, A., 2015. Analytical methods in untargeted metabolomics: state of the art in 2015. *Frontiers in bioengineering and biotechnology*, 3(March), p.23. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4350445&tool=pmcentrez&rendertype=abstract>.
- Anon, 2007. CERMAX. Centro de Ressonância Magnética António Xavier. Available at: <http://cermax.itqb.unl.pt/> [Accessed September 5, 2016].
- Avila, M.A. et al., 1998. Regulation by Hypoxia of Methionine Adenosyltransferase Activity and Gene Expression in Rat Hepatocytes. , pp.364–371.
- Banoei, M.M. et al., 2014. Metabolomics in critical care medicine: a new approach to biomarker discovery. *Clinical and investigative medicine. Médecine clinique et experimentale*, 37(6), pp.E363-76. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25618269>.
- Barroso, H., Meliço-Silvestre, A. & Taveira, N., 2014. Microbiologia Médica - Volume 1. In *Microbiologia Médica*. Lisboa, pp. 316–331.
- Beckonert, O. et al., 2007. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc*, 2(11), pp.2692–2703. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18007604>.
- Bottai, D. et al., 2014. Mycobacterium tuberculosis Exploits Asparagine to Assimilate Nitrogen and Resist Acid Stress during Infection. , 10(2).
- Boudah, S. et al., 2014. Annotation of the human serum metabolome by coupling three liquid chromatography methods to high-resolution mass spectrometry. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 966, pp.34–47.
- Brauer, R. et al., 2011. Preanalytical standardization of amino acid and acylcarnitine metabolite profiling in human blood using tandem mass spectrometry. *Metabolomics*, 7(3), pp.344–352.
- Brennan, L., 2014. NMR-based metabolomics: From sample preparation to applications

- in nutrition research. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 83, pp.42–49.
- Broa, R. & Smildeab, A.K., 2014. Principal component analysis. *Analytical methods*, pp.1–16.
- Chadzynski, G.L. et al., 2014. In vivo proton magnetic resonance spectroscopic imaging of the healthy human brain at 9.4 T: initial experience. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 28(3), pp.239–249. Available at: <http://link.springer.com/10.1007/s10334-014-0460-5>.
- Chen, Y. et al., 2013. ¹H-NMR Spectroscopy Revealed Mycobacterium tuberculosis Caused Abnormal Serum Metabolic Profile of Cattle. *PLoS ONE*, 8(9), pp.1–14.
- Cruz, T. et al., 2014. ¹H NMR Analysis of Cerebrospinal Fluid from Alzheimer's Disease Patients: An Example of a Possible Misinterpretation Due to Non-Adjustment of pH. *Metabolites*, 4(1), pp.115–28. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4018676&tool=pmcentrez&rendertype=abstract>.
- Daykin, C.A. et al., 2002. The comparison of plasma deproteinization methods for the detection of low-molecular-weight metabolites by (1)H nuclear magnetic resonance spectroscopy. *Analytical biochemistry*, 304(2), pp.220–30. Available at: http://www.researchgate.net/publication/11361178_The_Comparison_of_Plasma_Deproteinization_Methods_for_the_Detection_of_Low-Molecular-Weight_Metabolites_by_1H_Nuclear_Magnetic_Resonance_Spectroscopy.
- Deborde, C. et al., 2013. High-resolution ¹H-NMR spectroscopy and beyond to explore plant metabolome. *Advances in Botanical Research*, 67, pp.1–66. Available at: <http://www.sciencedirect.com/science/article/pii/B9780123979223000010>.
- Delogu, G., Sali, M. & Fadda, G., 2013. The biology of mycobacterium tuberculosis infection. *Mediterranean Journal of Hematology and Infectious Diseases*, 5(1).
- Devaux, P.G., Horning, M.G. & Horning, E.C., 1971. Benzyloxime derivatives of steroids. A new metabolic profile procedure for human urinary steroids human urinary steroids. *Analytical Letts*, 4(3), pp.151–160.
- Diaz, S.O. et al., 2011. Metabolic biomarkers of prenatal disorders: An exploratory NMR metabonomics study of second trimester maternal urine and blood plasma. *Journal of Proteome Research*, 10(8), pp.3732–3742.

- Diniz, A. et al., 2015. Infecção por VIH, SIDA e Tuberculose em números – 2015. *Direcção Geral de Saúde*, p.72.
- Dunn, W.B. et al., 2014. Molecular phenotyping of a UK population: defining the human serum metabolome. *Metabolomics*, 11(1), pp.9–26. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4289517&tool=pmcentrez&rendertype=abstract%5Cnhttp://link.springer.com/10.1007/s11306-014-0707-1>.
- Dunn, W.B. et al., 2011. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature protocols*, 6(7), pp.1060–1083.
- Duque-correa, M.A. et al., 2014. Macrophage arginase-1 controls bacterial growth and pathology in hypoxic tuberculosis granulomas.
- Emwas, A.H. et al., 2016. Recommendations and Standardization of Biomarker Quantification Using NMR-Based Metabolomics with Particular Focus on Urinary Analysis. *Journal of Proteome Research*, 15(2), pp.360–373.
- Emwas, A.H.M. et al., 2013. NMR-based metabolomics in human disease diagnosis: Applications, limitations, and recommendations. *Metabolomics*, 9(5), pp.1048–1072.
- Fan, T.W.M., 1996. Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 28(2), pp.161–219.
- Fend, R. et al., 2006. Prospects for clinical application of electronic-nose technology to early detection of Mycobacterium tuberculosis in culture and sputum. *Journal of Clinical Microbiology*, 44(6), pp.2039–2045.
- Ferreira, M.M.C. et al., 1999. Quimiometria i: Calibração multivariada, um tutorial. *Quimica Nova*, 22(5), pp.724–731.
- Finkelstein, J.D., 1990. Methionine metabolism in mammals. , 1.
- Fogel, N., 2015. Tuberculosis: A disease without boundaries. *Tuberculosis*, 95(5), pp.5–9. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1472979214206950>.
- Frediani, J.K. et al., 2014. Plasma metabolomics in human pulmonary tuberculosis disease: A pilot study. *PLoS ONE*, 9(10).
- García-rodríguez, J.F. et al., 2011. Extrapulmonary tuberculosis : epidemiology and risk

- factors. , 29(7), pp.502–509.
- Gil, V.M.. & Geraldes, C.F.G.C., 2002. *Ressonância Magnética Nuclear Fundamentos, Métodos e Aplicações* 2nd ed.,
- Gromski, P.S. et al., 2015. A tutorial review: Metabolomics and partial least squares-discriminant analysis - a marriage of convenience or a shotgun wedding. *Analytica Chimica Acta*, 879, pp.10–23. Available at: <http://dx.doi.org/10.1016/j.aca.2015.02.012>.
- Guthrie, J.W., 2012. *General considerations when dealing with biological fluid samples*, Elsevier. Available at: <http://dx.doi.org/10.1016/B978-0-12-381373-2.10065-1>.
- Haas, C.T. et al., 2016. Diagnostic “omics” for active tuberculosis. *BMC Medicine*, 14(1), p.37. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27005907><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4804573><http://www.biomedcentral.com/1741-7015/14/37>.
- Heather, L.C. et al., 2013. A practical guide to metabolomic profiling as a discovery tool for human heart disease. *Journal of Molecular and Cellular Cardiology*, 55(1), pp.2–11.
- Issaq, H.J. et al., 2009. Analytical and statistical approaches to metabolomics research. *Journal of Separation Science*, 32(13), pp.2183–2199.
- JPC, S., AM, R. & F, E., 2000. *Mycobacterium*. In *Microbiologia* F. WFC & S. JCF, eds., Lisboa: Lidel - Edições Técnicas.
- Kaddurah-Daouk, R., Kristal, B.S. & Weinshilboum, R.M., 2008. Metabolomics: a global biochemical approach to drug response and disease. *Annual review of pharmacology and toxicology*, 48, pp.653–83. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18184107>.
- Kaufmann, S.H.E. & Parida, S.K., 2008. Tuberculosis in Africa: Learning from Pathogenesis for Biomarker Identification. *Cell Host and Microbe*, 4(3), pp.219–228.
- Khan, A.R. et al., 2011. NMR spectroscopy based metabolic profiling of urine and serum for investigation of physiological perturbations during radiation sickness. *Metabolomics*, 7(4), pp.583–592.
- Kremer, L. et al., 2002. The M. tuberculosis antigen 85 complex and mycolyltransferase

- activity. *Letters in Applied Microbiology*, 34(4), pp.233–237.
- L. Eriksson, et al., 2006. *Multi - and megavariate data analysis* 2nd ed. Umetrics: Umea, ed., Sweden.
- Lau, S.K.P. et al., 2015. Metabolomic profiling of plasma from patients with tuberculosis by use of untargeted mass spectrometry reveals novel biomarkers for diagnosis. *Journal of Clinical Microbiology*, 53(12), pp.3750–3759. Available at: <http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L607012915%5Cnhttp://dx.doi.org/10.1128/JCM.01568-15%5Cnhttp://sfx.library.uu.nl/utrecht?sid=EMBASE&issn=1098660X&id=doi:10.1128/JCM.01568-15&atitle=Metabolomic+profiling+of+plasma+from+>.
- Lawn, S.D. & Zumla, A.I., 2011. Tuberculosis. *The Lancet*, 378(9785), pp.57–72.
- Leibfritz, D., Dreher, W. & Willker, W., 2007. The Handbook of Metabonomics and Metabolomics. *The Handbook of Metabonomics and Metabolomics*, pp.489–516. Available at: <http://www.sciencedirect.com/science/article/pii/B9780444528414500187>.
- Lenz, E.M. & Wilson, I.D., 2007. Analytical strategies in metabonomics. *Journal of Proteome Research*, 6(2), pp.443–458.
- Lesch, J.E., 1996. Species and specificity. An interpretation of the history of immunology - Mazumdar,PMH. *Science*, 273(5271), pp.75–76.
- Leung, G. et al., 2015. Lower Risk of Tuberculosis in Obesity. , 167, pp.1297–1304.
- Levine, J. et al., 2000. Stability of CSF metabolites measured by proton NMR. *Journal of Neural Transmission*, 107(7), pp.843–848.
- Lewandowski, C.M., Co-investigator, N. & Lewandowski, C.M., 2015. WHO Global tuberculosis report 2015. *The effects of brief mindfulness intervention on acute pain experience: An examination of individual difference*, 1, pp.1689–1699.
- Ma, Y. et al., 2012. Metabolomics in the fields of oncology: A review of recent research. *Molecular Biology Reports*, 39(7), pp.7505–7511.
- Meena, R. & Meena, L.S., 2011. Unique Characteristic Features of Mycobacterium Tuberculosis in Relation to Immune System Rajni and Laxman S . Meena Institute of Genomics and Integrative Biology , Allergy and Infectious Diseases ,. *American Journal of Immunology*, 7(1), pp.1–8.
- Mickiewicz, B. et al., 2014. Metabolic profiling of serum samples by 1H nuclear magnetic

- resonance spectroscopy as a potential diagnostic approach for septic shock. *Critical care medicine*, 42(5), pp.1140–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24368342>.
- Mittelstrass, K. et al., 2011. Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genetics*, 7(8).
- Moore, S.C. et al., 2014. Human metabolic correlates of body mass index. *Metabolomics : Official journal of the Metabolomic Society*, 10(2), pp.259–269. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4169991&tool=pmcentrez&rendertype=abstract>.
- Murray, P.R., Rosenthal, K.S. & Pfaller, M.A., 2011. *Microbiologia Médica* 6^a., Rio de Janeiro.
- Nayak, S. & Acharjya, B., 2012. VDRL test and its interpretation. *Indian Journal of Dermatology*, 57(1), p.3. Available at: <http://www.e-ijd.org/text.asp?2012/57/1/3/92666>.
- Nicholson, J.K., Buckingham, M.J. & Sadler, P.J., 1983. High resolution 1 H n.m.r. studies of vertebrate blood and plasma. *Biochemical Journal*, 211(3), pp.605–615. Available at: <http://www.biochemj.org/content/211/3/605.abstract>.
- Nicholson, J.K., Holmes, E. & Lindon, J.C., 2007. Metabonomics and metabolomics techniques and their applications in mammalian system. In *The Handbook of Metabonomics and Metabolomics*. pp. 1–33.
- Nicholson, J.K., Timbrell, J.A. & Sadler, P.J., 1985. Proton NMR spectra of urine as indicators of renal damage. Mercury-induced nephrotoxicity in rats. *Molecular Pharmacology*, 27(6), pp.644–651. Available at: <http://www.scopus.com/inward/record.url?eid=2-s2.0-0022254437&partnerID=tZOtx3y1>.
- Nielsen, J. & Oliver, S., 2005. The next wave in metabolome analysis. *Trends in Biotechnology*, 23(11), pp.544–546.
- Oliver, S., 2003. Functional Genomics: All the King's Horses AND All the King's Men CAN Put Humpty Together Again. *Molecular Cell*, 12(6), pp.1343–1344.
- van Oss, C.J., 1989. On the mechanism of the cold ethanol precipitation method of plasma protein fractionation. *Journal of Protein Chemistry*, 8(5), pp.661–668.
- Pai, M., Riley, L.W. & Colford, J.M., 2004. Interferon-?? assays in the immunodiagnosis

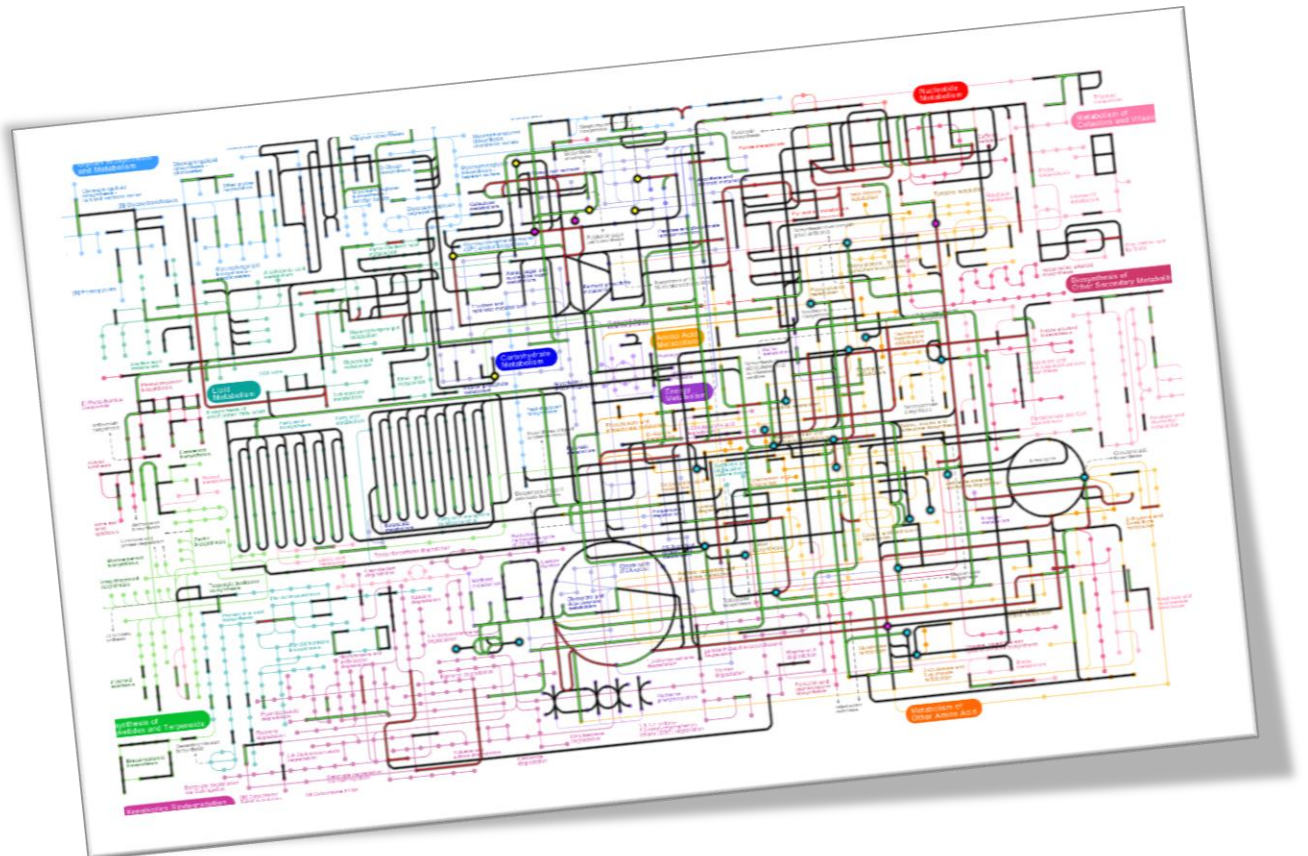
- of tuberculosis: A systematic review. *Lancet Infectious Diseases*, 4(12), pp.761–766.
- Paper, O., 2015. Evaluation of Pulmonary and Extrapulmonary Tuberculosis in Immunocompetent Adults : A Retrospective Case Series Analysis. , pp.75–79.
- Pavlou, A.K. et al., 2004. Detection of Mycobacterium tuberculosis (TB) in vitro and in situ using an electronic nose in combination with a neural network system. *Biosensors and Bioelectronics*, 20(3), pp.538–544.
- Polson, C. et al., 2003. Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography-tandem mass spectrometry. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 785(2), pp.263–275.
- Powers, R., 2009. NMR metabolomics and drug discovery. *Magnetic resonance in chemistry : MRC*, 47 Suppl 1(June), pp.S2–S11.
- du Preez, I. & Loots, D.T., 2014. Can metabolomics improve tuberculosis diagnostics? *Metabolomics*, pp.1–10.
- Psychogios, N., Hau, D.D., et al., 2011. The human serum metabolome. *PLoS ONE*, 6(2), p.430.
- Psychogios, N., Hau, D.D., et al., 2011. The human serum metabolome. *PloS one*, 6(2), p.e16957. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3040193&tool=pmcentrez&rendertype=abstract>.
- Rolin, D. et al., 2013. High-resolution 1H-NMR spectroscopy and beyond to explore plant metabolome. *Advances in Botanical Research*, 67, pp.1–66.
- Sasindran, S.J. & Torrelles, J.B., 2011. Mycobacterium tuberculosis infection and inflammation: What is beneficial for the host and for the bacterium? *Frontiers in Microbiology*, 2(JAN), pp.1–16.
- Sharma, U. et al., 2015. Metabolic abnormalities of gastrointestinal mucosa in celiac disease: An *in vitro* proton nuclear magnetic resonance spectroscopy study. *Journal of Gastroenterology and Hepatology*, 30(10), pp.1492–1498. Available at: <http://doi.wiley.com/10.1111/jgh.12979>.
- Shin, J. et al., 2011. H NMR-based Metabolomic Profiling in Mice Infected with Mycobacterium tuberculosis. , pp.2238–2247.

- Smith, I., 2003. *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clinical microbiology reviews*, 16(3), pp.463–496. Available at: <http://cmr.asm.org/content/16/3/463.short>.
- Smolinska, A. et al., 2012. NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. *Analytica Chimica Acta*, 750, pp.82–97.
- Somashekar, B.S. et al., 2012. Metabolomic signatures in guinea pigs infected with epidemic-associated W-Beijing strains of mycobacterium tuberculosis. *Journal of Proteome Research*, 11(10), pp.4873–4884.
- Somashekar, B.S. et al., 2012. Tuberculosis.
- Sumner, L.W., Mendes, P. & Dixon, R.A., 2003. Plant metabolomics: Large-scale phytochemistry in the functional genomics era. *Phytochemistry*, 62(6), pp.817–836.
- Sunnetcioglu, A. et al., 2015. Comparative analysis of pulmonary and extrapulmonary tuberculosis of 411 cases. *Annals of Clinical Microbiology and Antimicrobials*, 14(1), p.34. Available at: <http://www.ann-clinmicrob.com/content/14/1/34>.
- Teahan, O. et al., 2006. Impact of Analytical Bias in Metabonomic Studies of Human Blood Serum and Plasma. , 78(13), pp.4307–4318.
- The Economist, 2014. The final frontier. , pp.2010–2012. Available at: <http://www.economist.com/news/asia/21607837-fixing-dreadful-sanitation-india-requires-not-just-building-lavatories-also-changing>.
- Tiziani, S. et al., 2008. Optimized metabolite extraction from blood serum for 1H nuclear magnetic resonance spectroscopy. *Analytical Biochemistry*, 377(1), pp.16–23. Available at: <http://www.scopus.com/inward/record.url?eid=2-s2.0-43049140446&partnerID=40&md5=5db54624fc5c6bbc8f145092be5f84c8>.
- Transfer, C. & Restraint, M., 1970. Experimental methodology.
- Triba, M.N. et al., 2015. PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol. BioSyst.*, 11(1), pp.13–19. Available at: <http://xlink.rsc.org/?DOI=C4MB00414K>.
- Trimigno, A. et al., 2015. Definition of food quality by NMR-based foodomics. *Current Opinion in Food Science*, 4, pp.99–104.
- Ulrich, E.L. et al., 2008. BioMagResBank. *Nucleic Acids Research*, 36(SUPPL. 1), pp.402–408.

- Umetrics, 2011. *A user guide to SIMCA-P*,
- Via, L.E. et al., 2008. Tuberculous Granulomas Are Hypoxic in Guinea Pigs , Rabbits , and Nonhuman Primates □. , 76(6), pp.2333–2340.
- Viant, M.R., 2003. Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochemical and Biophysical Research Communications*, 310(3), pp.943–948.
- Viant, M.R., Rosenblum, E.S. & Tjeerdema, R.S., 2003. NMR-Based Metabolomics: A Powerful Approach for Characterizing the Effects of Environmental Stressors on Organism Health. *Environmental Science and Technology*, 37(21), pp.4982–4989.
- Villas-B??as, S.G. et al., 2006. *Metabolome Analysis: An Introduction*,
- Voet, D. & Voet, J.G., 1996. Amino acid metabolism. In *Biochemistry*. pp. 727–776.
- Vuckovic, D., 2012. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. *Analytical and Bioanalytical Chemistry*, 403(6), pp.1523–1548.
- Vuckovic, D., 2013. *Sample Preparation in Global Metabolomics of Biological Fluids and Tissues*, Elsevier. Available at: <http://linkinghub.elsevier.com/retrieve/pii/B9780123944467000042>.
- Wedge, D.C. et al., 2011. Is serum or plasma more appropriate for intersubject comparisons in metabolomic studies? An assessment in patients with small-cell lung cancer. *Analytical Chemistry*, 83(17), pp.6689–6697.
- Weiner, J. et al., 2012. Biomarkers of Inflammation, Immunosuppression and Stress with Active Disease Are Revealed by Metabolomic Profiling of Tuberculosis Patients. *PLoS ONE*, 7(7), p.e40221. Available at: <http://dx.plos.org/10.1371/journal.pone.0040221>.
- Weljie, A.M. et al., 2006. Targeted Profiling : Quantitative Analysis of 1 H NMR Metabolomics Data. , 78(13), pp.4430–4442.
- Wilson, D.M. et al., 1974. Deuterium and carbon-13 tracer studies of ethanol metabolism in the rat by 2H, 1H-decoupled 13C nuclear magnetic resonance. *Biochemical and Biophysical Research Communications*, 56(3), pp.828–835.
- Winning, H. et al., 2008. Quantitative analysis of NMR spectra with chemometrics. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 190(1), pp.26–32. Available at:

- <http://www.sciencedirect.com/science/article/pii/S1090780707003242>.
- Wishart, D.S. et al., 2013. HMDB 3.0-The Human Metabolome Database in 2013. *Nucleic Acids Research*, 41(D1), pp.801–807.
- Wishart, D.S., 2008. Metabolomics: applications to food science and nutrition research. *Trends in Food Science and Technology*, 19(9), pp.482–493.
- Xia, J. et al., 2016. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. *Current Protocols in Bioinformatics*, (September), p.14.10.1-14.10.91. Available at: <http://doi.wiley.com/10.1002/cpbi.11>.
- Xia, J. & Wishart, D.S., 2011. *Metabolomic data processing, analysis, and interpretation using MetaboAnalyst*,
- Xia, J. & Wishart, D.S., 2010. MSEA: A web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Research*, 38(SUPPL. 2), pp.71–77.
- Yu, Z. et al., 2011. Differences between human plasma and serum metabolite profiles. *PLoS ONE*, 6(7), pp.1–6.
- Yu, Z. et al., 2012. Human serum metabolic profiles are age dependent. *Aging Cell*, 11(6), pp.960–967.
- Zhang, A. et al., 2012. Modern analytical techniques in metabolomics analysis. *The Analyst*, 137(2), p.293.
- Zhang, S. et al., 2008. Correlative and quantitative ¹H NMR-based metabolomics reveals specific metabolic pathway disturbances in diabetic rats. *Analytical Biochemistry*, 383(1), pp.76–84.
- Zhang, X. et al., 2013. Metabolic signatures of esophageal cancer: NMR-based metabolomics and UHPLC-based focused metabolomics of blood serum. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1832(8), pp.1207–1216. Available at: <http://dx.doi.org/10.1016/j.bbadis.2013.03.009>.
- Zhou, A. et al., 2013. Application of ¹H NMR spectroscopy-based metabolomics to sera of tuberculosis patients. *Journal of Proteome Research*, 12(10), pp.4642–4649.

Anexos



Anexo I

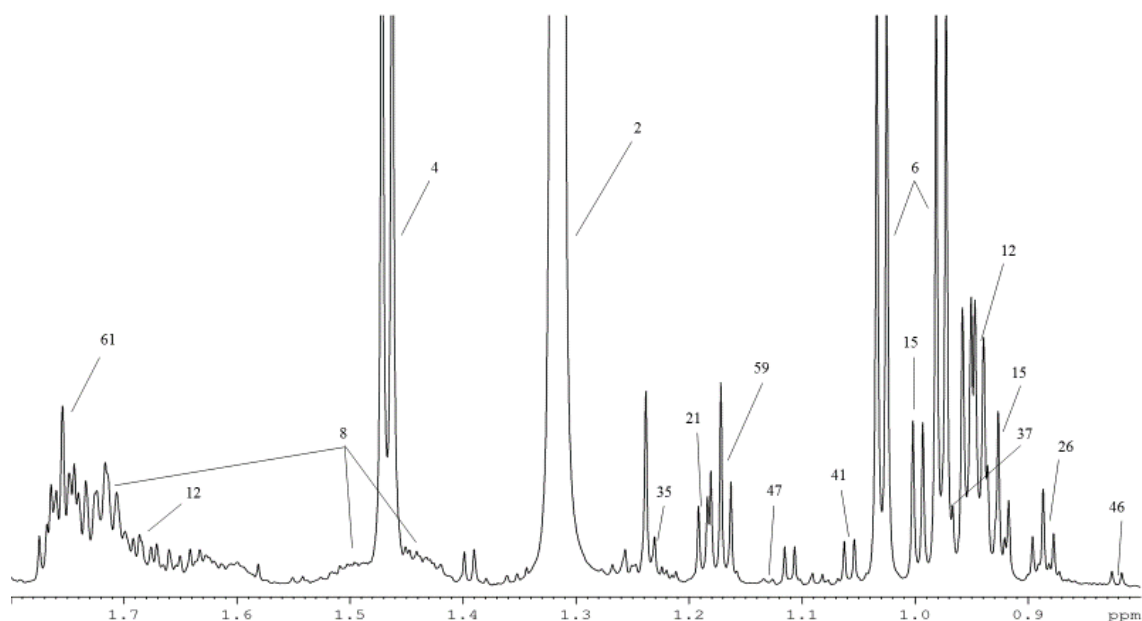


Figura A 1: Região espectral A, de uma amostra controle, entre os 0.8 e os 1.8 ppm. 2: lactato; 4: alanina; 6: valina; 8: lisina; 12: leucina; 15: isoleucina; 21: 3-hidroxibutirato; 26: 2-hidroxibutirato; 35: fucose; 37: 2-aminobutirato; 46: 2-hidroxisovalerato; 47: propilenoglicol; 59: etanol; 61: DSS. Ampliação 64x

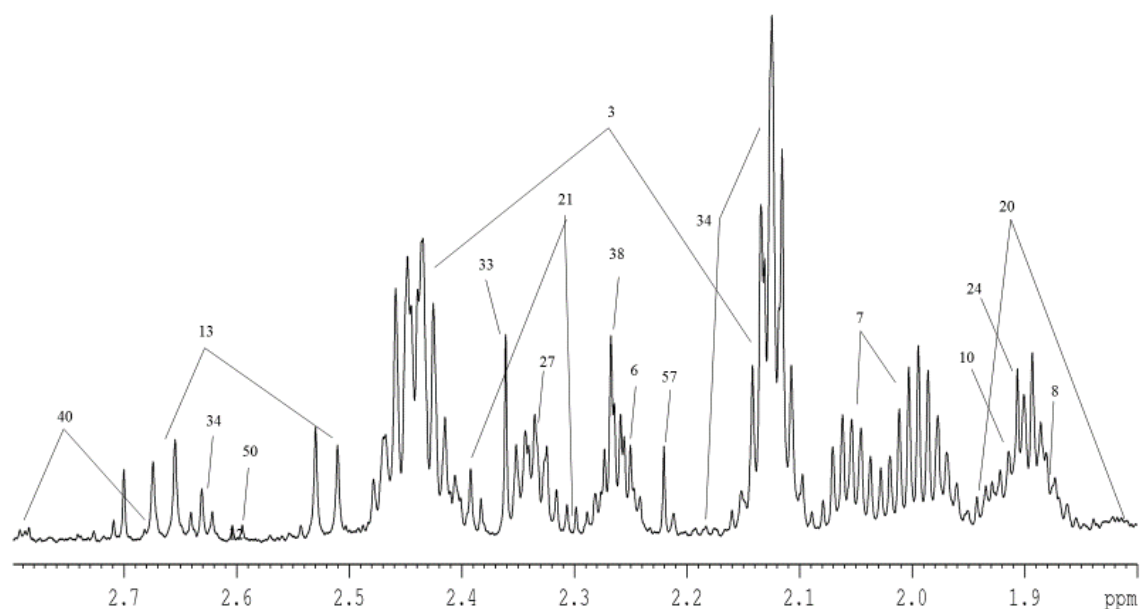


Figura A 2: Região B entre os 1.8 e os 2.8 ppm de um espectro de uma amostra controle. 3: glutamina; 6: valina; 7: prolina; 8: lisina; 10: arginina; 13: citrato; 20: ornitina; 21: 3-hidroxibutirato; 24: acetato; 27: glutamato; 33: piruvato; 34: metionina; 38: acetoacetato; 40: aspartato; 50: 2-oxoisocaproato; 57: acetona. Ampliação 64x

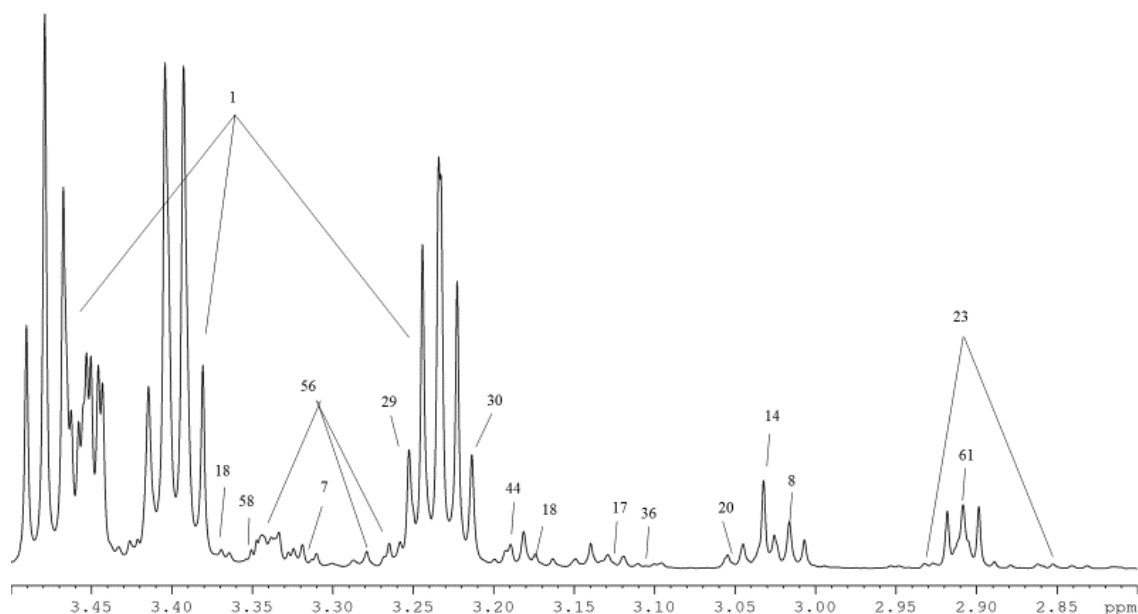


Figura A 3: Região C entre os 2.8 e os 3.5 ppm de um espectro de uma amostra controle. 1: glucose; 7: prolina; 8: lisina; 14: creatinina; 17: histidina; 18: cistina; 20: ornitina; 23: asparagina; 29: betaína; 30: carnitina; 36: cisteína; 44: colina; 56: 1,5-anidro sorbitol; 58: cafeína; 61: DSS. Ampliação 16x

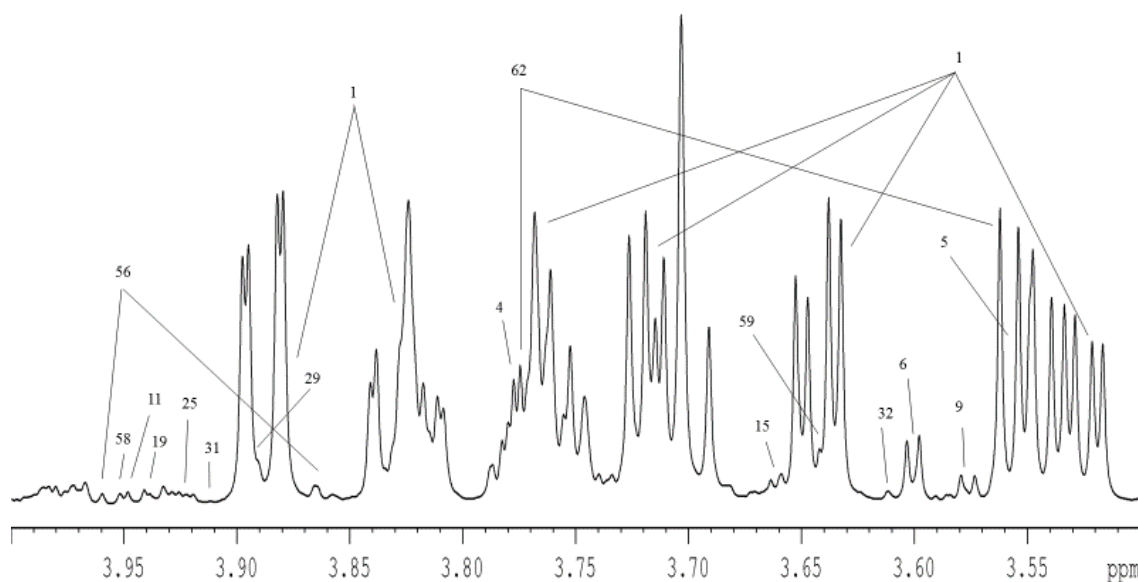


Figura A 4: Região D entre os 3.5 e os 4 ppm de um espectro de uma amostra controle. 1: glucose; 4: alanina; 5: glicina; 6: valina; 9: treonina; 11: serina; 15: isoleucina; 19: tirosina; 25: manose; 29: betaína; 31: creatina; 32: mio-inositol; 56: 1,5-anidro sorbitol; 58: cafeína; 59: etanol; 62: glicerol Ampliação 8x

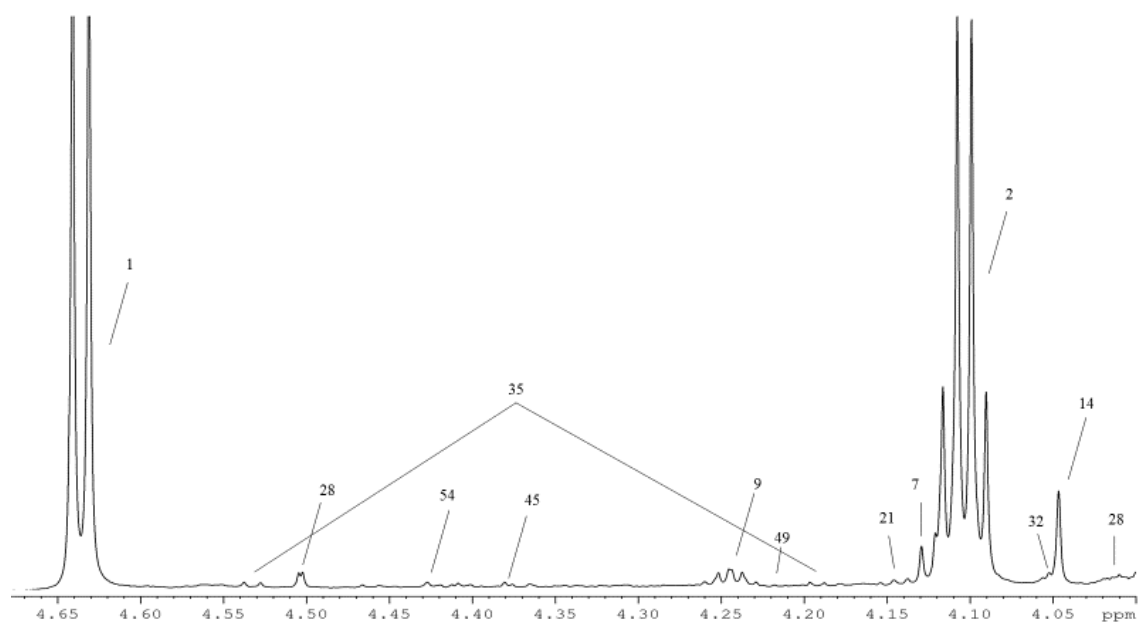


Figura A 5: Região E entre os 4 e os 4,7 ppm de um espectro de uma amostra controle. 1: glucose; 2: lactato; 7: prolina; 9: treonina; 14: creatinina; 21: 3-hidroxibutirato; 28: ascorbato; 35: fucose; 45: trans-4-hidroxi-L-prolina; 49: uridina; 54: trigonelina. Ampliação 32x

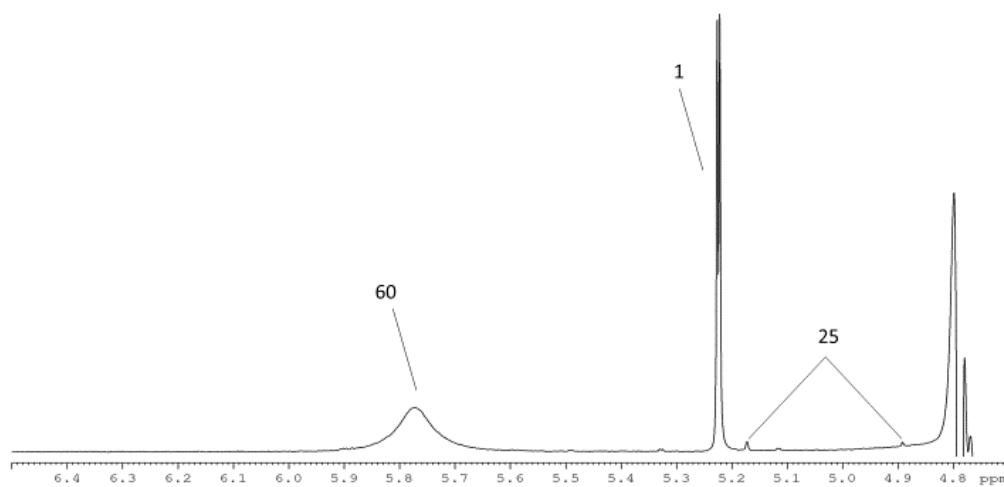


Figura A 6: Região F entre os 4,7 e os 6,5 ppm de um espectro de uma amostra controle. 1: glucose; 25: manose; 60: ureia. Ampliação 32x

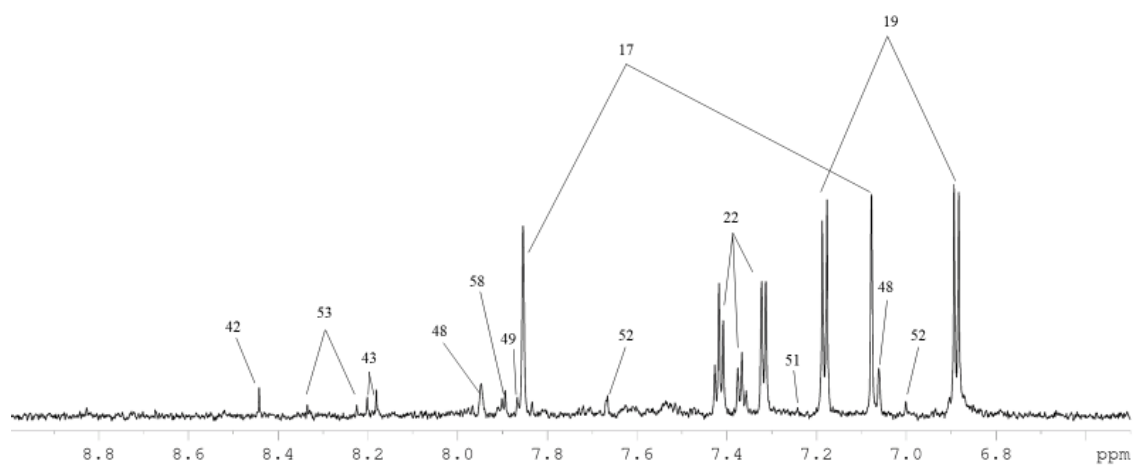


Figura A7: Região G entre os 6,5 e os 9 ppm de um espectro de uma amostra controlo. 17: histidina; 19: tirosina; 22: fenilalanina; 42: formato; 43: hipoxantina; 48: π -metilhistidina; 49: uridina; 51: triptofano; 52: τ -metilhistidina; 53: inosina. Ampliação 32x

Anexo II

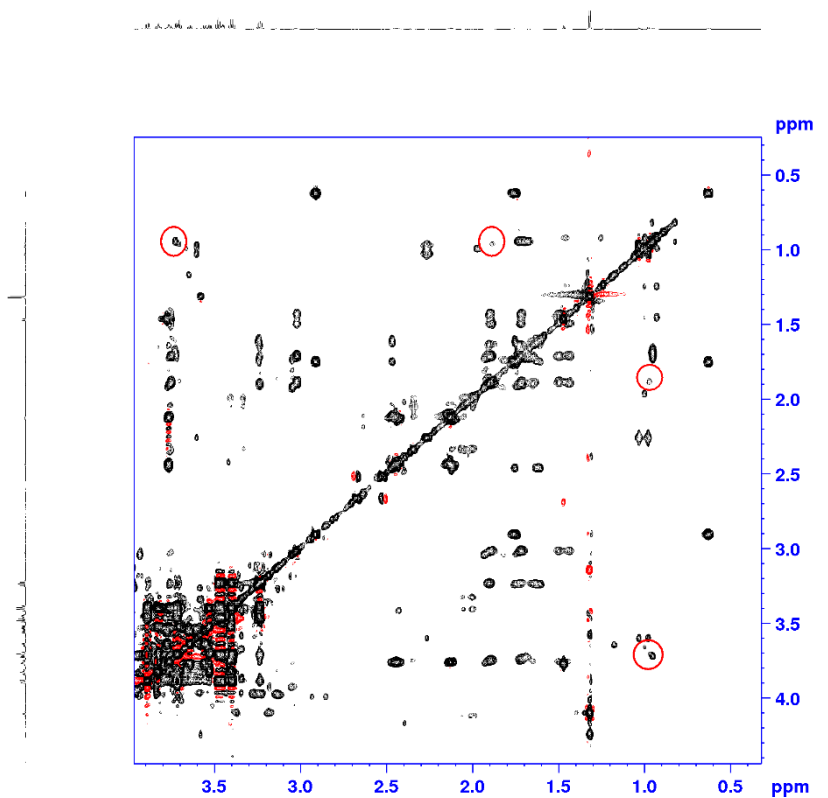


Figura B 1: região do espectro TOCSY utilizada para a identificação do 2-aminobutirato

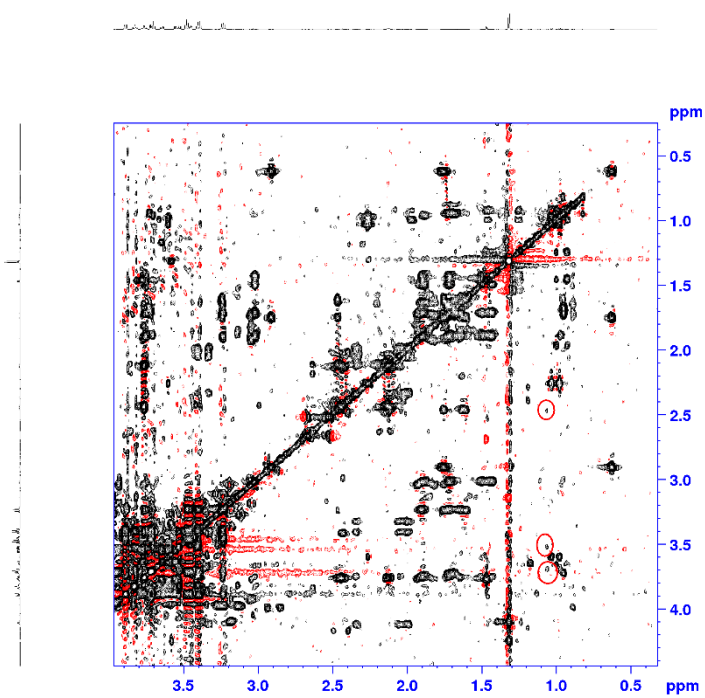


Figura B 2: região do espectro TOCSY utilizada para a identificação do 3-hidroxisobutirato

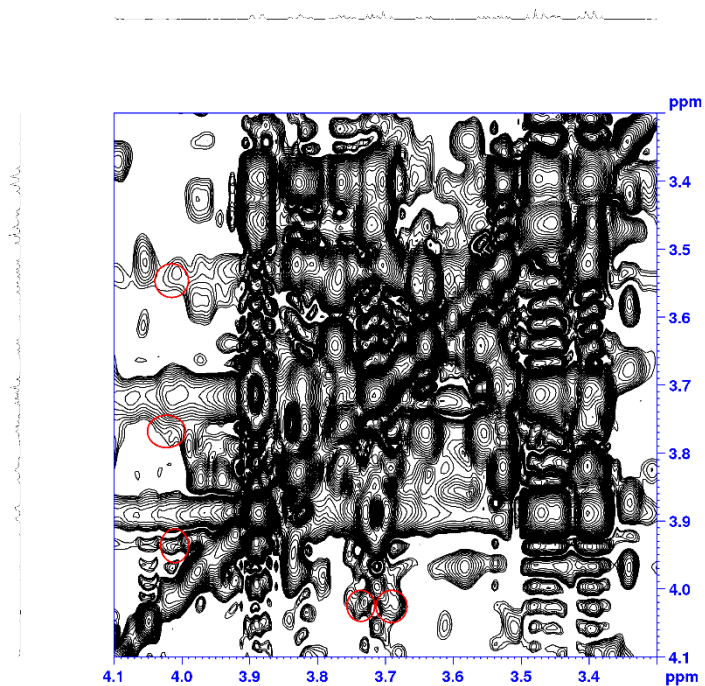


Figura B 3: região do espectro TOCSY utilizada para a identificação da frutose

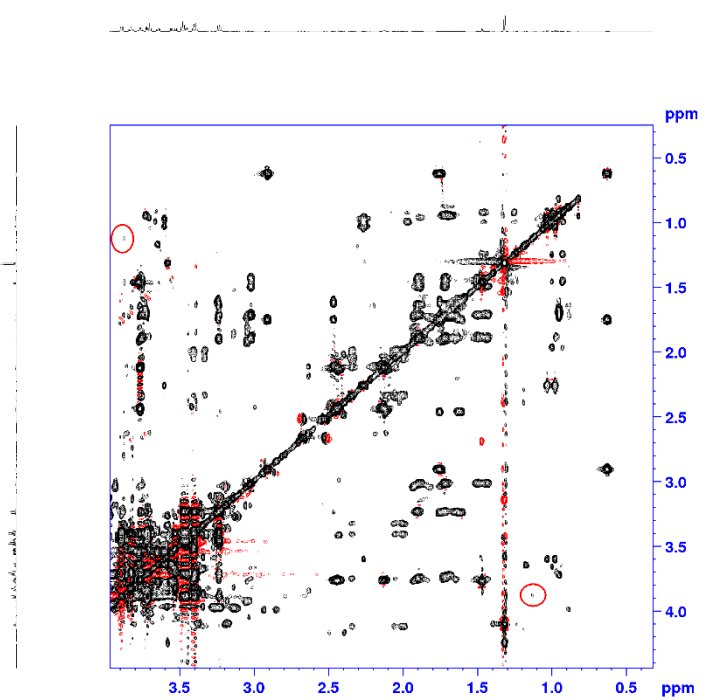


Figura B 3: região do espectro TOCSY utilizada para a identificação do propileno glicol

Anexo III

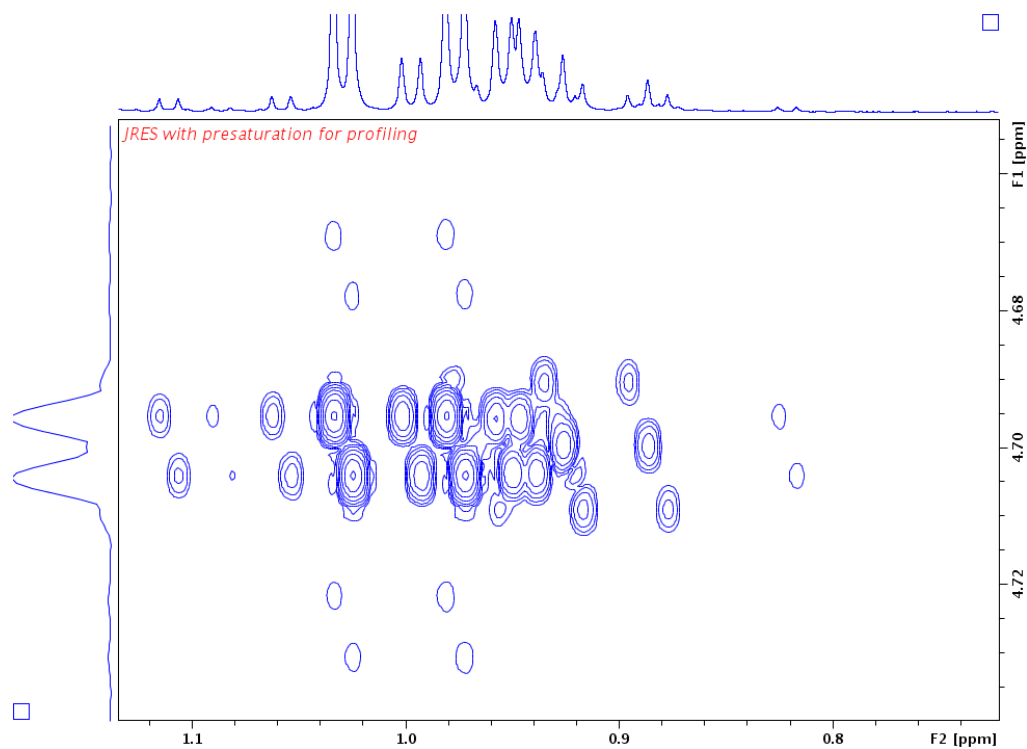


Figura C 1: região do espectro J-resolved de 0.74 a 1.14 ppm

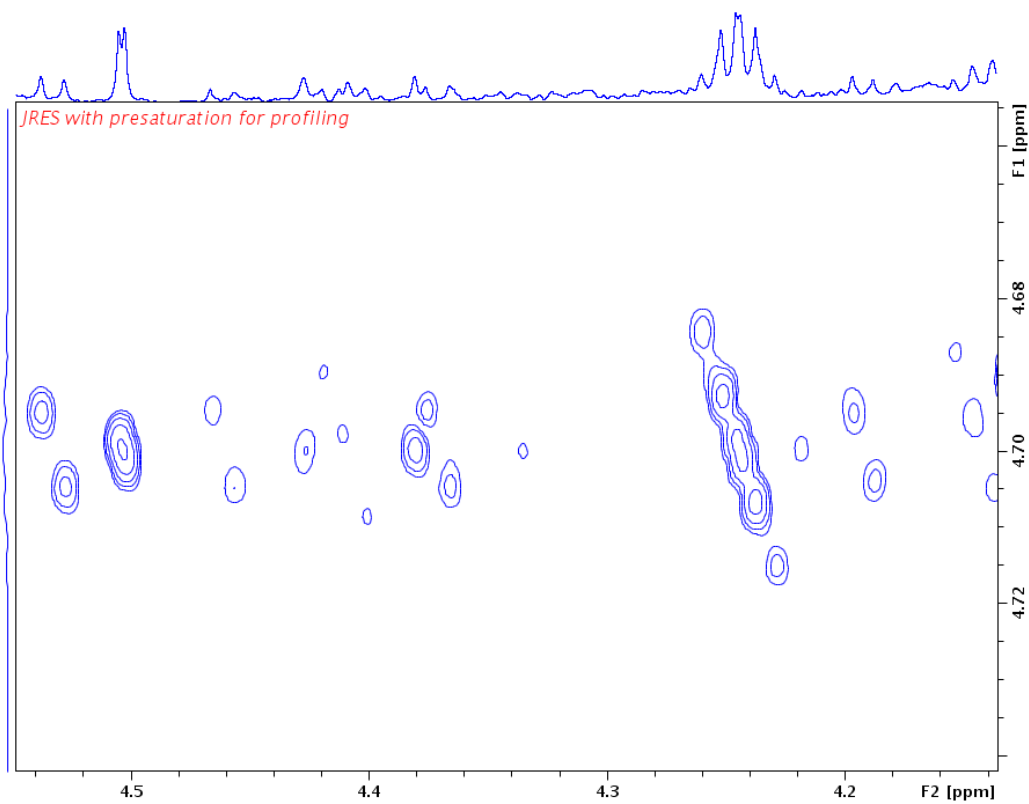


Figura C 2: região do espectro J-resolved de 4.14 a 4.54 ppm

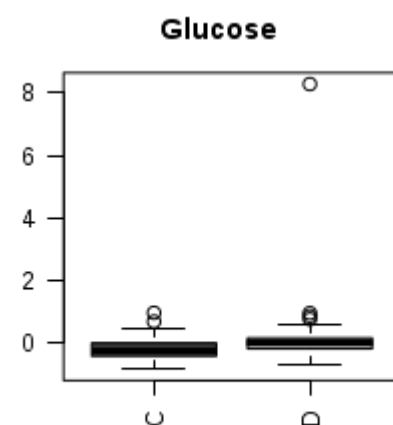
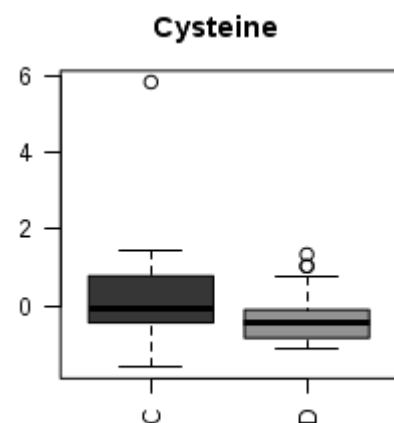
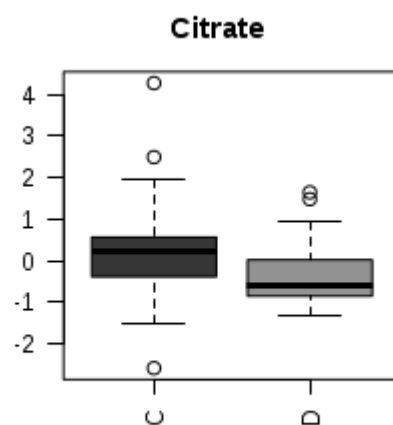
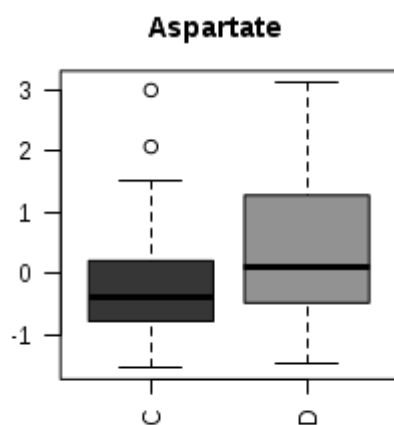
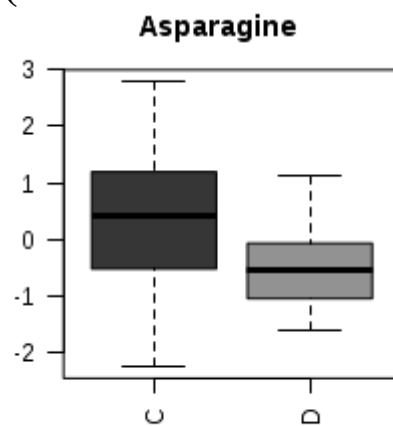
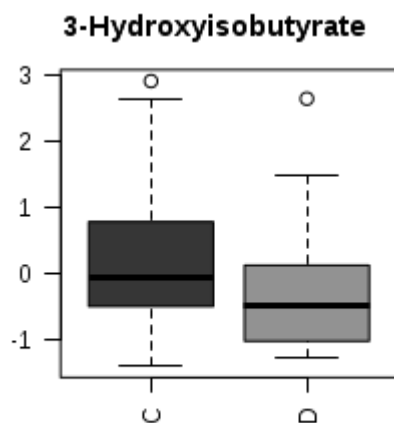
Anexo IV

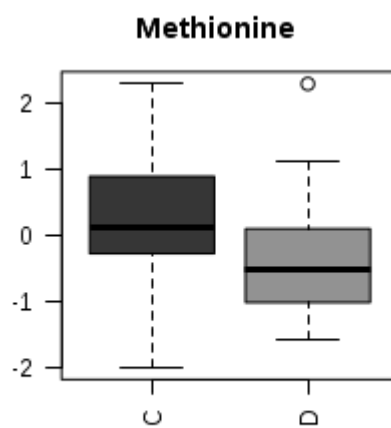
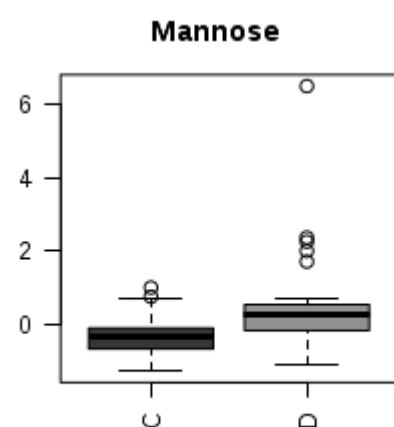
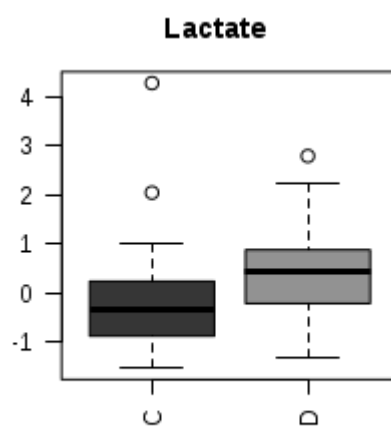
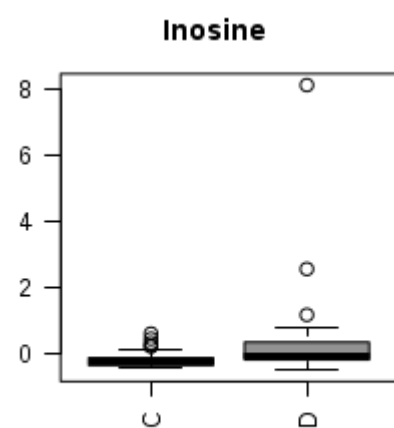
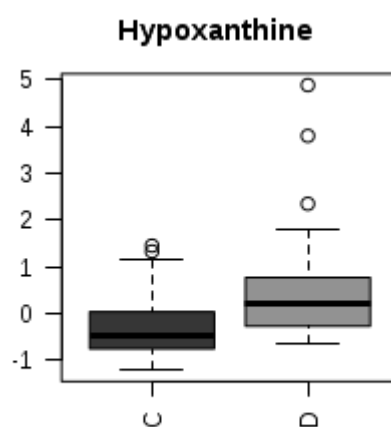
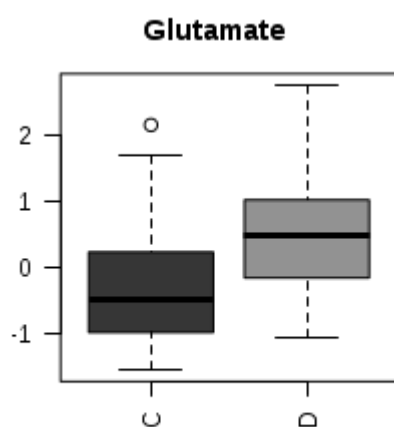
Tabela 18: Metabolitos identificados ordenados por classes, e respetivo desvio químico (δ). S: singlete; d: dupla; dd: duas dupletas; t: tripleto; q: quarteto; m: multipeto . Os sinais a negrito foram usados ara fazer o *assignment* dos metabolitos.

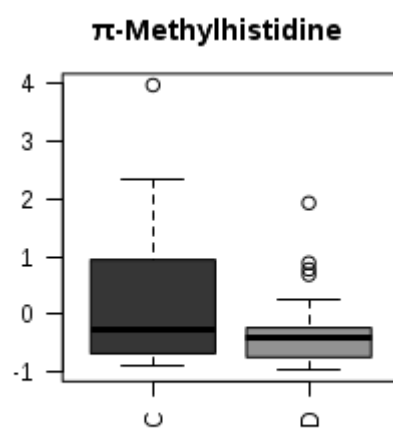
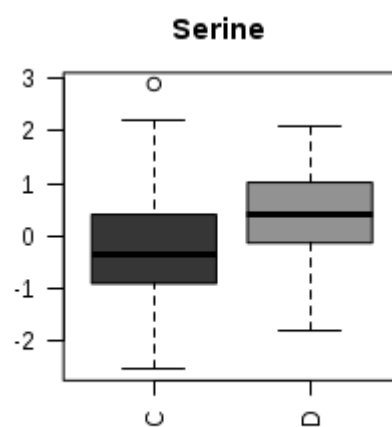
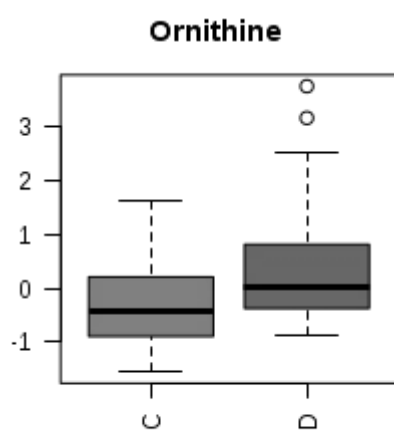
Classes	Metabolitos	δ ^1H (ppm)/ multiplicidade
Álcoois	Etanol	1.17 (t); 3.65 (q)
	Glicerol	3.56/3.64 (dd); 3.77 (dd)
	Propileno glicol	1.13 (d); 3.44; 3.54; 3.88
Aminoácidos	Alanina	1.47 (d); 3.77 (q)
	Arginina	1.64 (m); 1.72 (m) ; 1.89 (m); 1.92(m); 3.24(t); 3.76; 6.67; 7.22
	Asparagina	2.85(dd); 2.94 (dd); 3.99 (dd); 6.91; 7.62
	Aspartato	2.66(dd); 2.80 (dd); 3.89 (dd)
	Betaína	3.25 (m); 3.89 (s)
	Cisteína	3.02; 3.11 (m); 3.99 (dd)
	Cistina	3.18 (dd); 3.38 (dd); 4.10 (dd)
	Fenilalanina	3.11 (dd); 3.27 (dd); 3.99 (dd); 7.32 (m); 7.37 (m); 7.42 (m)
	Glicina	3.55 (s)
	Glutamato	2.04(m); 2.12 (m); 2.33(m); 3.75 (t)
	Glutamina	2.14 (m); 2.42 (m); 2.46; 3,76 (t); 6.88; 7.60
	Histidina	3.14 (dd); 3.24 (dd); 3.97 (dd); 7.08 (m); 7.85 (d)
	Isoleucina	0.93 (t); 1.00 (d); 1.25 (m); 1.46 (m); 1.97 (m); 3.66 (m)
	Leucina	0.94 (d); 0.95 (d); 1.67 (m); 1.74 (m); 3.73 (t)
	Lisina	1.47 (m); 1.72(m); 1.91 (m); 3.02 (t); 3.76 (t)
	Metionina	2.11 (s); 2.12 (m); 2.19 (m); 2.63 (t); 3.86 (dd)
	Ornitina	1.74 (m); 1.82 (m); 1.93 (m); 3.05 (t); 3.78 (t)
	Prolina	1.98 (m); 2.01 (m); 2.06 (m); 2.34; 3.32; 3.42; 4.12
	Serina	3.84 (dd); 3.94 (dd); 3.98 (dd)
	Tirosina	3.04 (dd); 3.19 (dd); 3.94 (dd); 6.89 (d); 7.18 (d)
	Trans-4-hidroxi- prolina	2.14; 2.42; 3.36; 3.47; 4.38; 4.66
	Treonina	1.32 (d); 3.58 (d); 4.25 (m)
	Triptofano	3.30 (dd); 3.47 (dd); 4.05 (dd); 7.19 (t); 7.27 (t); 7.31 (s); 7.54 (d); 7.72 (d); 10.18

Ácidos Orgânicos	Valina	0.98 (d); 1.03 (d); 2.26 (m); 3.60 (d)
	π (1)-metilhistidina	3.22; 3.30; 3.72; 3.94; 7.06 (s); 7.95 (s)
	τ (3)-metilhistidina	3.06; 3.16; 3.68; 3.96; 7.00 (s); 7.67 (s)
	2-aminobutirato	0.96 (t); 1.88; 3.70
	2-hidroxibutirato	0.88 (t); 1.64 (m); 1.73 (m); 3.98 (dd)
	2-hidroxisovalerato	0.82 (d); 0.95; 2.00; 3.84
	2-oxoisocaproato	0.92; 2.08; 2.60 (d)
	3-hidroxibutirato	1.19 (d); 2.29 (dd); 2.39 (dd); 4.14
	3-hidroxisobutirato	1.06 (d); 2.48; 3.53; 3.69
	Acetato	1.91 (s)
	Acetoacetato	2.27 (s); 3.42 (s)
	Ascorbato	3.72; 3.75; 4.01; 4.50 (d)
	Carnitina	2.41; 2.45 (m); 3.21 (s); 3.40; 3.43 (m); 4.56
	Citrato	2.52 (d); 2.66 (d)
	Formato	8.44 (s)
	Lactato	1.32 (d); 4.10 (q)
	Piruvato	2.36 (s)
	Malonato	3.13 (s)
Glícidos	Glucose	3.23 (t); 3.39 (t); 3.40 (t); 3.45 (m); 3.48 (t); 3.52 (dd); 3.70 (t); 3.71 (dd); 3.76 (dd); 3.82 (m); 3.83 (m); 3.89 (m); 4.64 (d); 5.22 (d)
	Frutose	3.55; 3.56; 3.59; 3.70; 3.79; 3.82; 3.89; 3.99; 4.01; 4.10 (m); 4.11 (d)
	Fucose	1.21; 1.22; 1.23 (s); 1.24 (s); 3.44; 3.56; 3.63; 3.74; 3.76; 3.79; 3.80; 3.85; 3.86; 3.87; 3.97; 4.01; 4.03; 4.07; 4.19; 4.53 (d); 5.19; 5.22; 5.27
	Manose	3.37; 3.55; 3.65; 3.73; 3.76; 3.80; 3.84; 3.87; 3.90; 3.93; 4.89 (d); 5.17 (d)
	Mio-inositol	3.27 (t); 3.53 (dd); 3.61 (t); 4.05 (t)
	1,5-anidro sorbitol	3.24; 3.26; 3.28; 3.34; 3.43; 3.68; 3.86; 3.88 (d); 3.97 (d)
Intermediários nucleosídeos	Hipoxantina	8.18 (s); 8.20 (s)
	Inosina	3.84; 3.91; 4.27; 4.43; 4.76 (t); 6.09 (d); 8.22 (s); 8.34 (s)
	Uridina	3.80; 3.90; 4.12; 4.22; 4.34; 5.89/ 5.90 (d); 7.86 (d)
Outros	Acetona	2.22 (s)
	Cafeína	3.35; 3.53; 3.95 (s); 7.89 (s)
	Colina	3.19 (s); 3.51 (m); 4.06 (m)
	Creatina	3.02 (s); 3.92 (s)
	Creatinina	3.03 (s); 4.05 (s)
	DSS	0.00 (s); 0.63; 1.76; 2.91
	Trigonelina	4.43 (s); 8.07; 8.82; 8.83; 9.11 (s)
	Ureia	5.78

Anexo V: Gráficos de bigodes (análise univariada)







Anexo VI

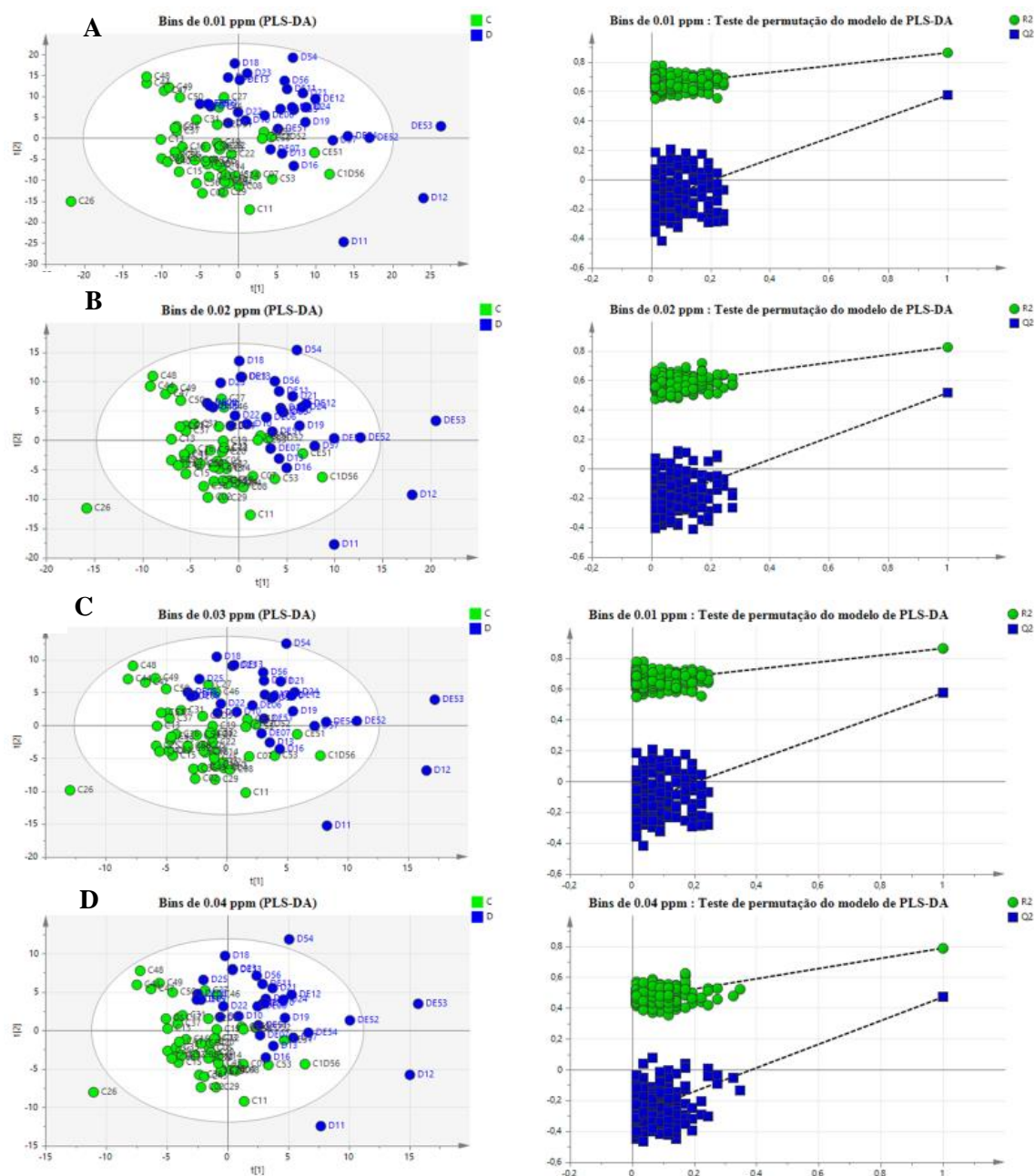


Figura D: Gráfico de *scores* do modelo de PLS-DA e respetivo teste de permutação utilizado para a escolha do tamanho ideal dos *bins*. Análises realizadas com *bins* de 0.01 ppm (A), 0.02 ppm (B), 0.03 ppm (C) e 0.04 ppm (D).

Anexo VII

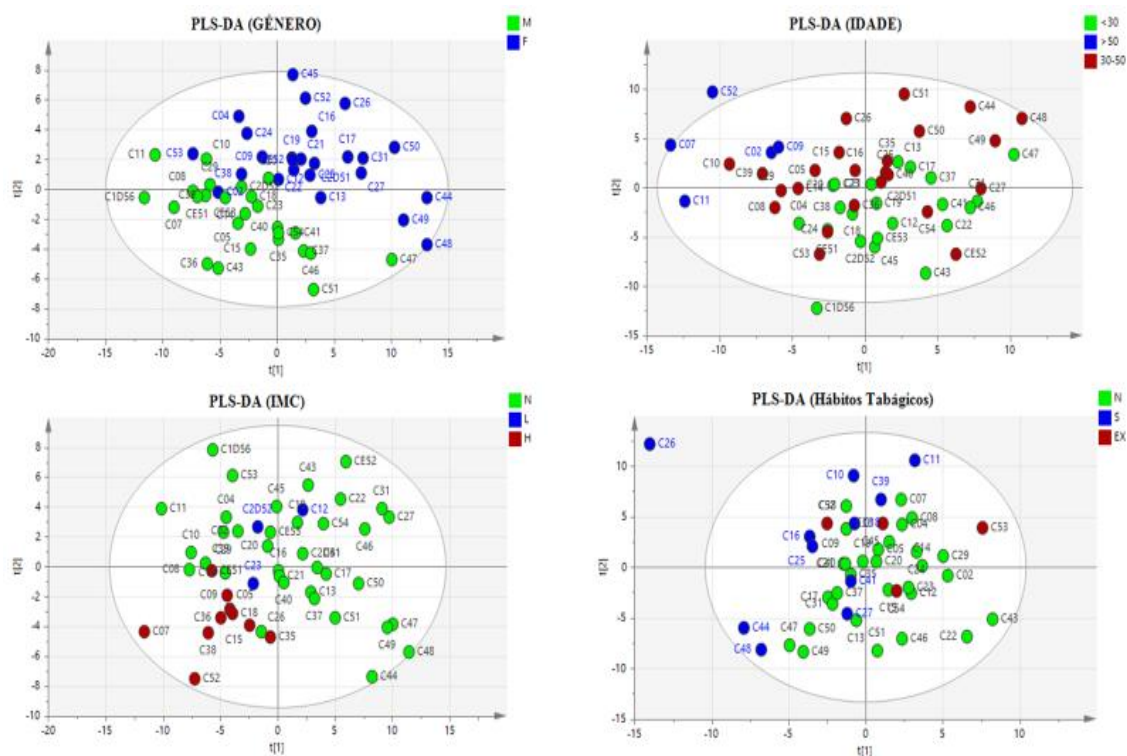


Figura E 1 1: Gráficos de *scores* dos modelos de PLS-DA das amostras controle, utilizados para analisar as variáveis externas através da análise *fingerprinting*. A: gênero; B: idade; C: IMC; D: Hábitos tabágicos

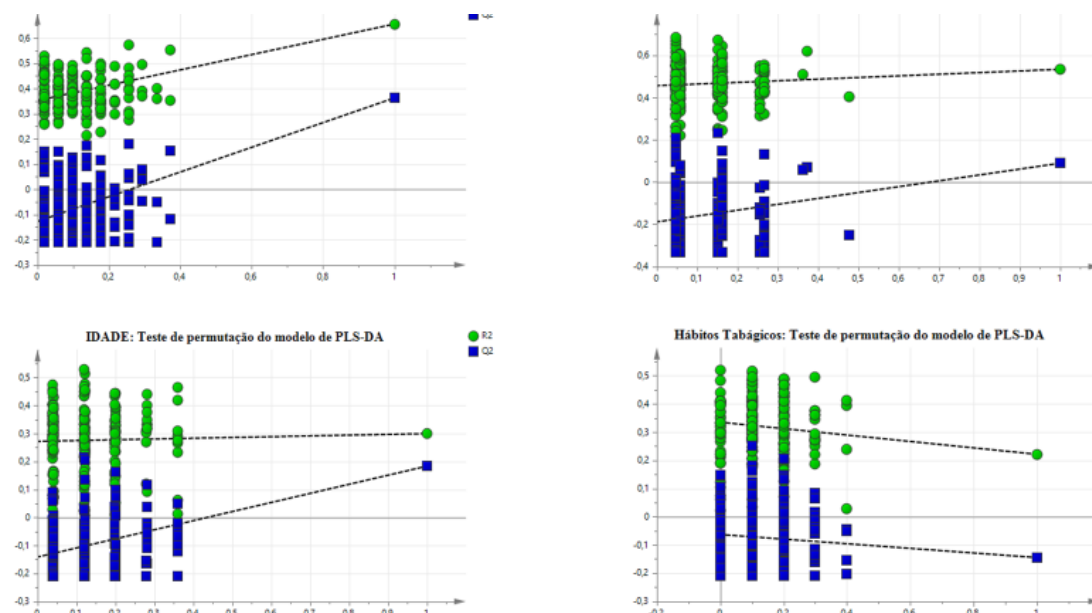


Figura E 1 2: Testes de permutação dos modelos de PLS-DA das amostras controle, utilizados para analisar as variáveis externas através da análise *fingerprinting*. A: gênero; B: idade; C: IMC; D: Hábitos tabágicos. Os pontos verdes representam os valores de R^2 e os pontos azuis representam os valores de Q^2

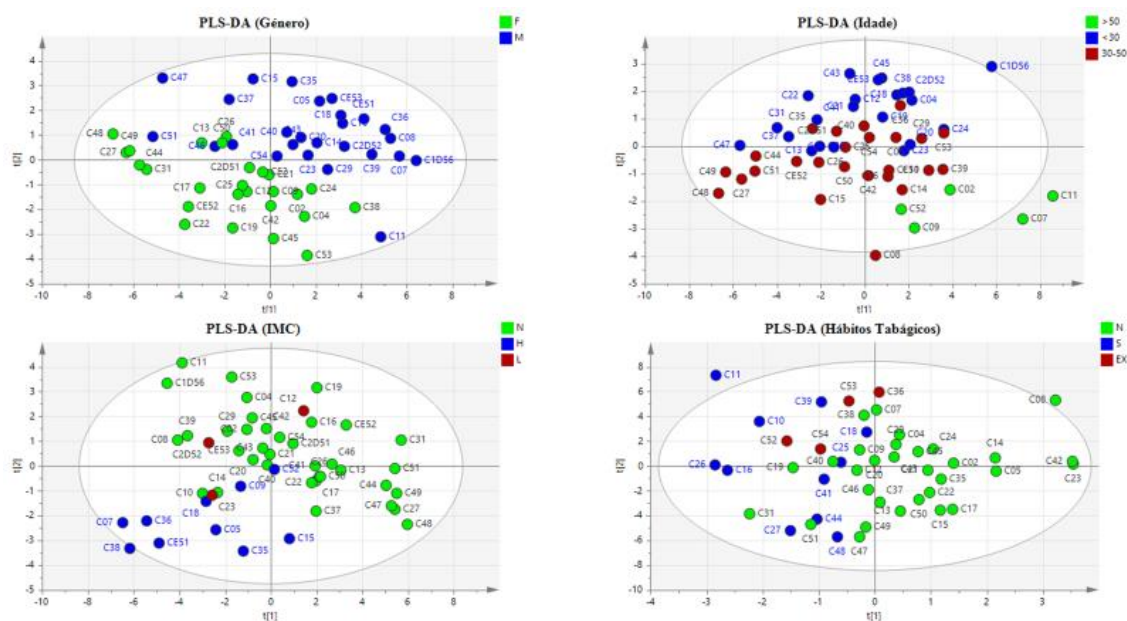


Figura E 1 3: Gráficos de *scores* dos modelos de PLS-DA das amostras controle, utilizados para analisar as variáveis externas através da análise do perfil metabólico. A: gênero; B: idade; C: IMC; D: Hábitos tabágicos

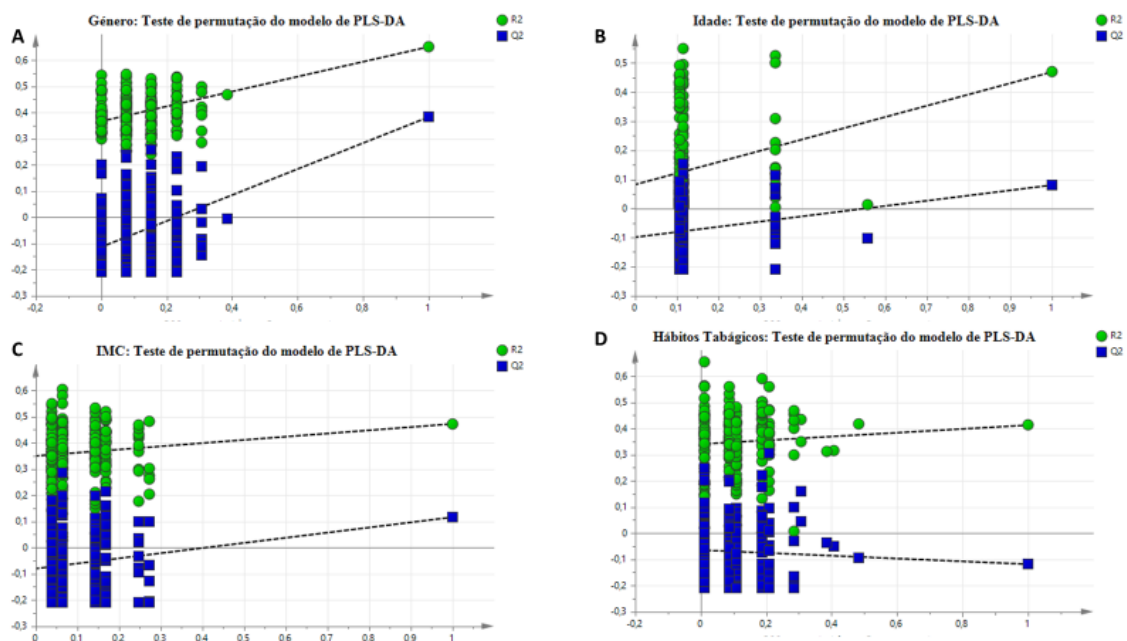


Figura E 1 4: Testes de permutação dos modelos de PLS-DA das amostras controle, utilizados para analisar as variáveis externas através da análise do perfil metabólico. A: gênero; B: idade; C: IMC; D: Hábitos tabágicos. Os pontos verdes representam os valores de R² e os pontos azuis representam os valores de Q²

Anexo VIII

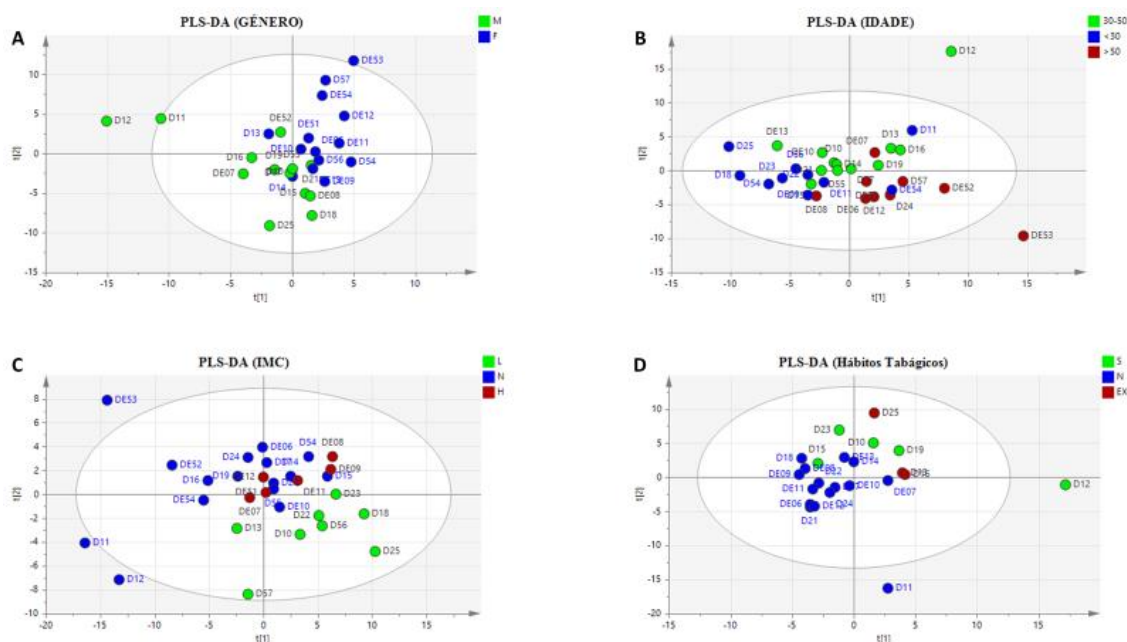


Figura F 1: Gráficos de *scores* dos modelos de PLS-DA das amostras de doentes, utilizados para analisar as variáveis externas através da análise *fingerprinting*. **A:** género; **B:** idade; **C:** IMC; **D:** Hábitos tabágicos

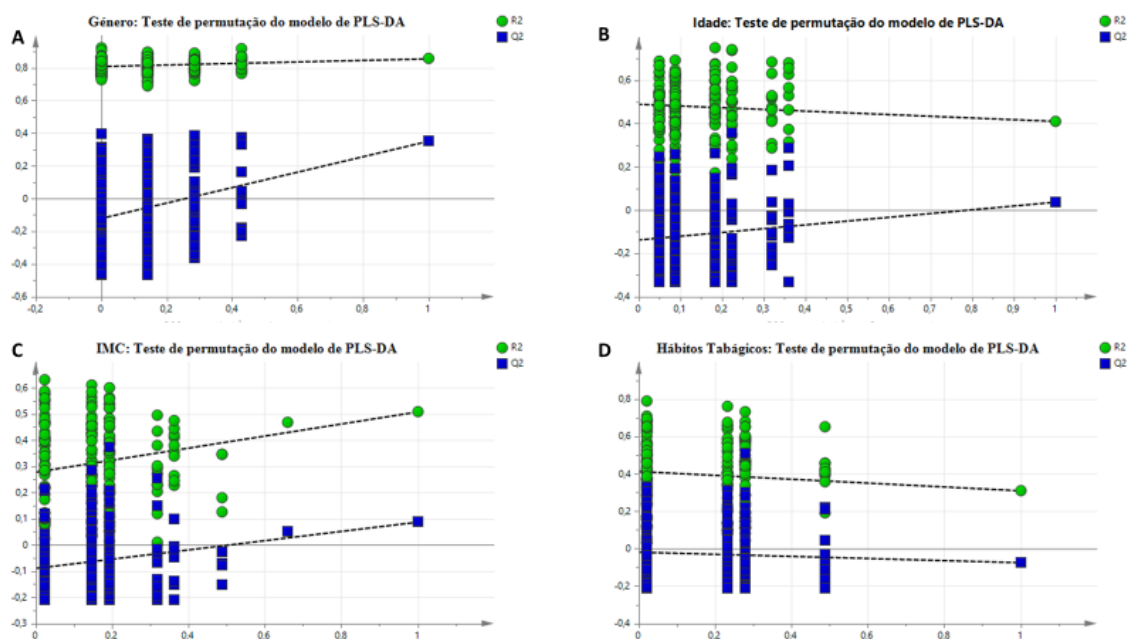


Figura F 2: Testes de permutação dos modelos de PLS-DA das amostras de doentes, utilizados para analisar as variáveis externas através da análise *fingerprinting*. **A:** género; **B:** idade; **C:** IMC; **D:** Hábitos tabágicos. Os pontos verdes representam os valores de R^2 e os pontos azuis representam os valores de Q^2

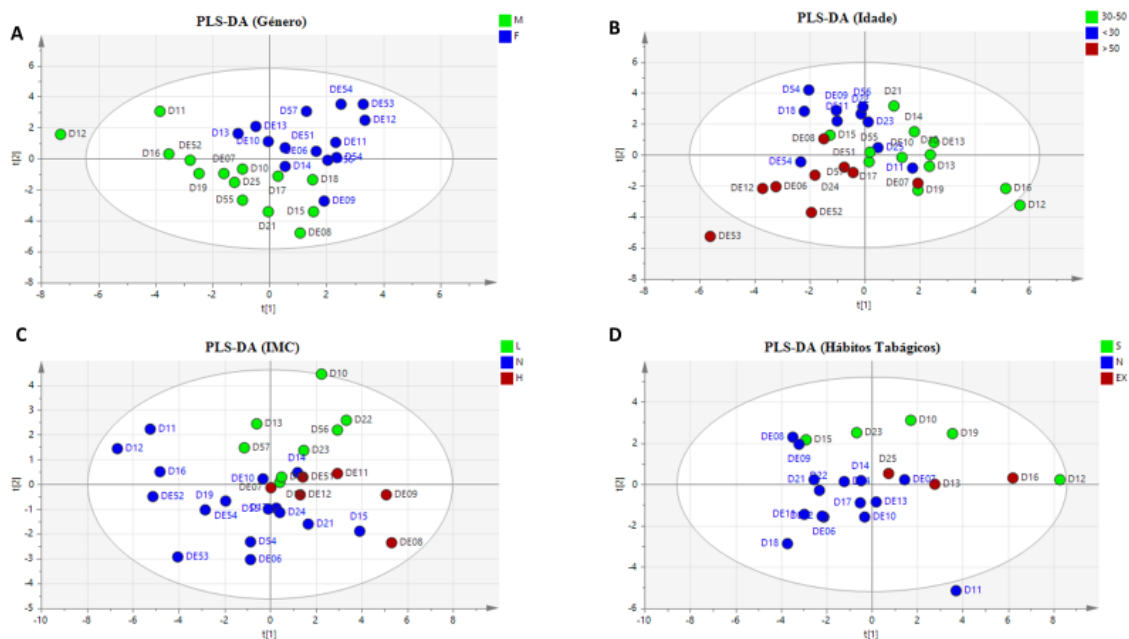


Figura F 3: Gráficos de *scores* dos modelos de PLS-DA das amostras de doentes, utilizados para analisar as variáveis externas através da análise do perfil metabólico. **A:** gênero; **B:** idade; **C:** IMC; **D:** Hábitos tabágicos

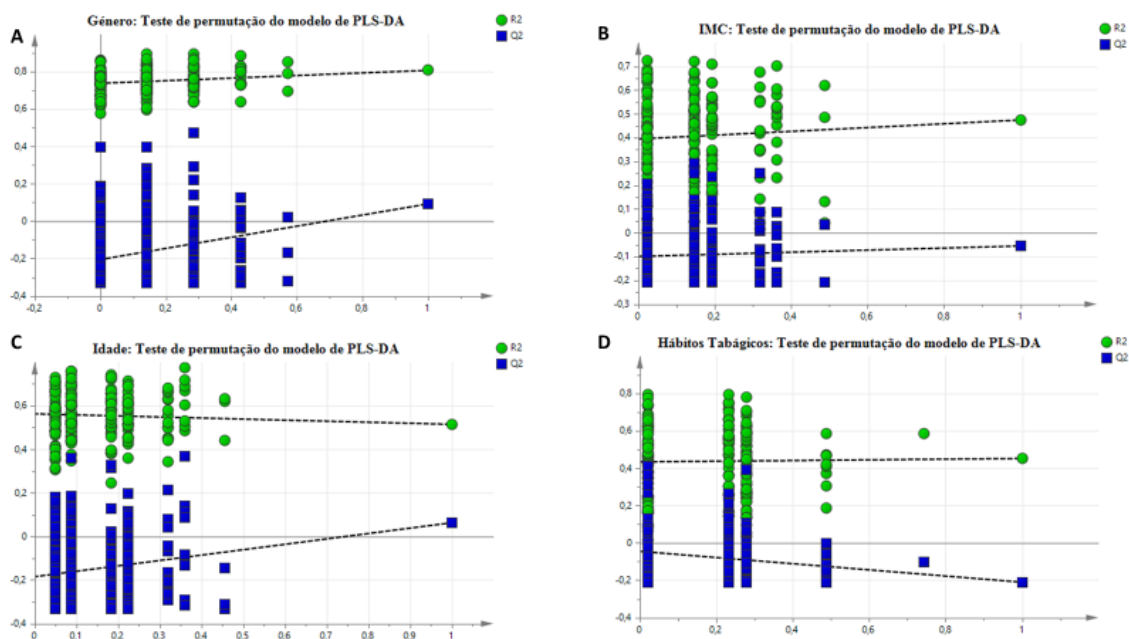


Figura F 4: Testes de permutação dos modelos de PLS-DA das amostras de doentes, utilizados para analisar as variáveis externas através da análise *fingerprinting*. **A:** gênero; **B:** idade; **C:** IMC; **D:** Hábitos tabágicos. Os pontos verdes representam os valores de R^2 e os pontos azuis representam os valores de Q^2

Anexo IX

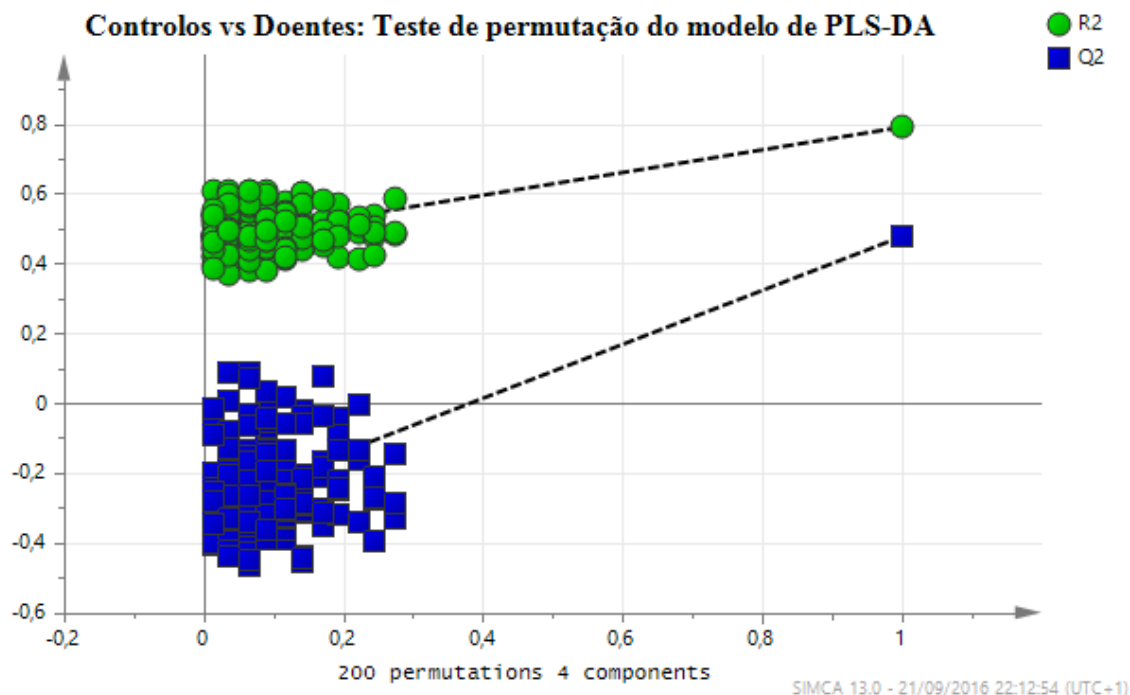


Figura G 1: Testes de permutação dos modelos de PLS-DA das amostras de controlos e de doentes, utilizados para analisar as variáveis externas através da análise *fingerprinting*. Os pontos verdes representam os valores de R^2 e os pontos azuis representam os valores de Q^2

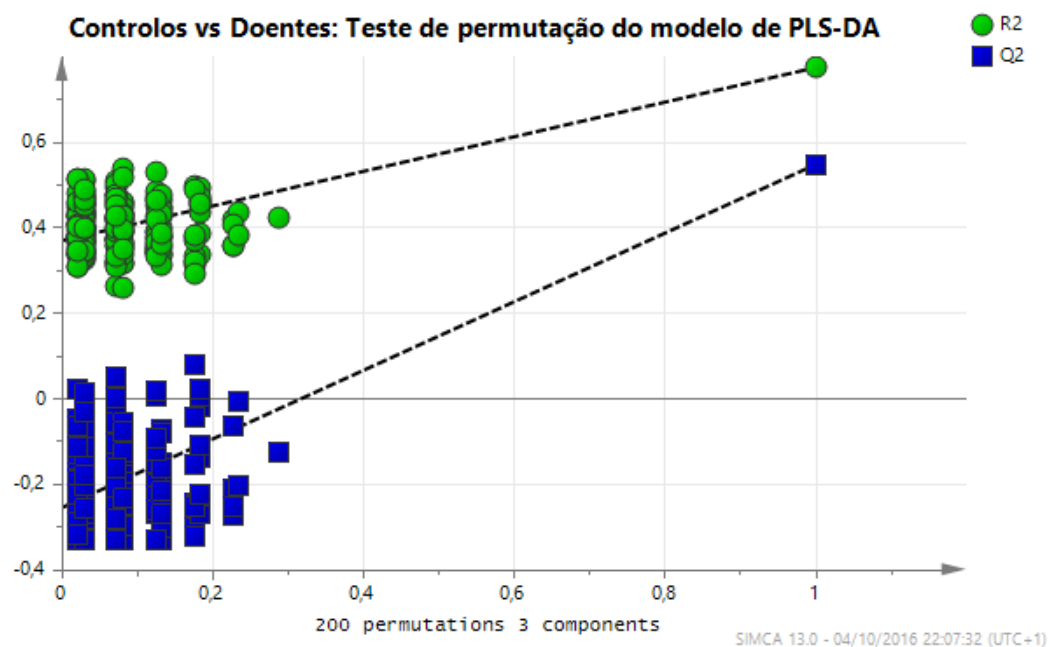


Figura G 2: Testes de permutação dos modelos de PLS-DA das amostras dos controlos e de doentes, utilizados para analisar as variáveis externas através da análise do perfil metabólico. Os pontos verdes representam os valores de R^2 e os pontos azuis representam os valores de Q^2

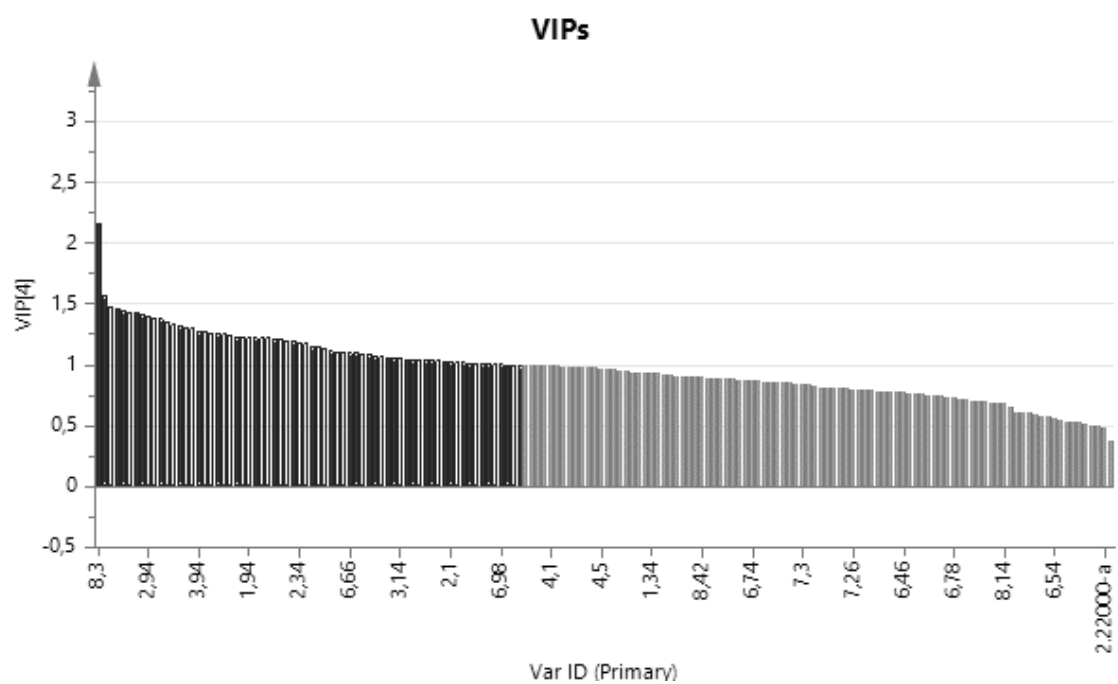


Figura G 3: Representação dos VIPs para o modelo de PLS-DA utilizando os dados da análise fingerprinting

